Lights, Camera,....Income!
Illuminating the National Accounts-Household Surveys Debate

Maxim Pinkovskiy* Xavier Sala-i-Martin
Federal Reserve Bank of New York Columbia University and NBER

December 30, 2015

Abstract

GDP per capita and household survey means present conflicting pictures of the rate of economic development in emerging countries. One of the areas in which the national accounts-household surveys debate is key is the measurement of developing world poverty. We propose a data-driven method to assess the relative quality of GDP per capita and survey means by comparing them to the evolution of satellite-recorded nighttime lights. Our main assumption, which is robust to a variety of specification checks, is that the measurement error in nighttime lights is unrelated to the measurement errors in either national accounts or survey means. We obtain estimates of weights on national accounts and survey means in an optimal proxy for true income; these weights are very large for national accounts and very modest for survey means. We conclusively reject the null hypothesis that the optimal weight on surveys is greater than the optimal weight on national accounts, and we generally fail to reject the null hypothesis that the optimal weight on surveys is zero. Additionally, we provide evidence that national accounts are good indicators of desirable outcomes for the poor (such as longer life expectancy, better education and access to safe water), and we show that surveys appear to perform worse in developing countries that are richer and that are growing faster. Therefore, we interpret our results as providing support for estimates of world poverty that are based on national accounts. JEL Codes: I32.

*Xavier Sala-i-Martin is the Jerome and Matthew Grossman Professor of Economics at Columbia University. We would like to thank Robert Barro, Lawrence Katz and four anonymous referees for excellent comments. We would like to thank Rajashri Chakrabarti, Richard Crump, Angus Deaton, Jerry Hausman, Marco Del Negro, Wilbert van der Klaauw, Donghoon Lee, Kyle Meng, Matthew Notowidigdo, Elias Papaioannou, Daria Pelech, Thomas Piketty, Miikka Rokkanen, Hannah Rubinton, Adam Sacarny, Louise Sheiner, Andrei Shleifer, Jon Steinsson, Giorgio Topa, Eric Verhoogen, Justin Wolfers, Basit Zafar, Christine Zhang and seminar participants at the Federal Reserve Bank of New York, the Brookings Institute and Columbia University for insightful and useful suggestions. We are very grateful to John Grigsby for superb research assistance. Any views expressed in this paper are the authors' and do not necessarily reflect those of the Federal Reserve Bank of New York or of the Federal Reserve System. All errors are our own.

1 Introduction

How much economic growth has there been in the developing world? The answer to this apparently simple question turns out to depend significantly on which dataset you use to measure economic growth. Much of the growth literature (e.g. the vast cross-country literature following Barro 1991) uses national accounts GDP per capita, but many researchers studying poverty and income distribution (Chen and Ravallion 2010; Lackner and Milanovic 2014) use household survey means. The problem is that for reasons that are not well understood (Deaton 2005), there is a large and growing discrepancy between living standards as measured by these two different sources. For example, if one looks at the national accounts, India's consumption per capita has grown by over 100% between 1994 and 2010, but if one looks at India's National Sample Survey (NSS), its consumption per capita increased by only 29% during this period. African countries show an even starker discrepancy. Angola's national accounts, for example, suggest that its consumption per capita has grown by 108% between 2000 and 2009. Angolan household surveys, on the other hand, show instead a 5% *decline* in consumption per capita. One area in which these discrepancies matter is the estimation of world poverty, with approaches based on household surveys delivering poverty rates five times as large as the approaches based on national accounts. Which of these radically divergent sets of numbers is right, and how can we tell?

So far, attempts to answer this question have mostly been theoretical. Proponents of national accounts, and proponents of household surveys have pointed out valid conceptual problems with each of these measures, but have not been able to decide which of these measures, or which combination of them, is the best to use. On the one hand, there is theoretical and empirical evidence that, as people grow richer, they are less likely to respond to surveys (Korinek et al. (2005)). This might explain why survey estimates of income and consumption might be too low and grow too slowly. On the other hand, national accounts measure living standards very indirectly, and often rely on outdated assumptions on the structure of the economy. For example, national accounts often measure agricultural output by multiplying acres under cultivation by a measure of agricultural productivity, which may be antiquated or updated in ways that have little to do with actual growth (Deaton 2005). Since no one has done a systematic empirical study that quantifies the potential sources of error line by line, it is still an open question whether national accounts or household surveys are better.

In this paper, we take a different approach to answering this question. Instead of trying to quantify the individual sources of error in the national accounts and household surveys, we show that if there were a third measure with independent measurement error, we could use it as a tool to see how much weight to give national accounts GDP per capita compared to survey means in measuring true income. We should emphasize that the crucial property of this third measure would not be that it is measured with little error, but that its measurement error would be independent of the measurement errors in GDP per capita and survey means. Luckily, such a measure exists...it is satellite-recorded data on nighttime lights from the surface of the Earth, which are visible from outer space (Henderson, Storeygard and Weil 2012). These nighttime lights are overwhelmingly generated by human activity (they come from buildings and cars in cities or on roads), and they have been shown to be correlated with economic activity (Elvidge et al. 1997, Henderson, Storeygard and Weil 2012, Chen and Nordhaus 2011, Michalopoulos and Papaioannou 2011, 2012 inter alia). Clearly, nighttime lights measure economic activity with error, but this error should have nothing to do with the nonresponse biases and faulty statistical assumptions that may plague national accounts and household surveys. The error in nighttime lights comes from climatic conditions such as auroral activity, cloudiness and humidity, or because of cultural attitudes towards lighting, all of which presumably are not associated with measurement errors in national accounts or household surveys. Unlike survey teams or statisical agencies, satellites collect nighttime lights data impersonally, and do not require compliance or truthfulness of the population generating the lights.

We can see intuitively how nighttime lights help us determine whether GDP per capita or survey means are better by returning to our example of India and Angola. We recall that according to household surveys, India's consumption per capita grew only by 29% between 1994 and 2010, but according to the national accounts its per capita consumption more than doubled during this period.[1] Figure I gives a view of India between 1994 and 2010. We see that lights in India increase dramatically both in their intensity over the major cities as well as in their extent over previously unlit areas of the country. In fact, the lights per capita measure increases by 112%, similar to the 100% increase in national accounts consumption per capita, and very different from the 29% increase in survey mean consumption.[2] Hence, the lights say that the growth rate as computed by GDP is likely to be close to the growth rate in true income. For both the lights and GDP per capita to overstate consumption growth would require a coincidence, because their measurement errors should be uncorrelated. Turning to the Angolan example, we recall that according to the household surveys, Angola has experienced a 5% *decline* in per capita income, while according to the national accounts, it has experienced a *doubling* of per capita income (108% growth) between 2000 and 2009. Figure II presents a picture of nighttime lights over southern Africa in 2000 and in 2009. We see that Angola has many more lights in 2009 than it did in 2000 (in fact, it experienced 103% growth in its lights per capita, almost exactly the same rate as the growth in GDP per capita). Again, the similar growth rates between

[1] For all statistics on levels and growth rates of national accounts GDP per capita, survey means and nighttime lights for all countries with survey data available in the period 1992-2010, see Online Appendix Table II

[2] We discuss how satellite photos of nighttime lights are converted into numerical light indices in Section 2, and we will define "lights per capita" in that section as well.

lights and national accounts could be a coincidence, but that would be a strange coincidence since the errors of these series are uncorrelated. Though their data generating processes are completely unrelated – the bad assumptions of statistical agencies should not make nighttime lights any brighter – national accounts and nighttime lights appear to move in tandem, which suggests that nighttime lights and national accounts are reflections of the same underlying true income concept.

To demonstrate formally that the examples of India and Angola are not coincidences, we use the same principle of asking whether nighttime lights closer track national accounts GDP per capita or survey means in order to estimate true income for a broad panel of developing countries. We compute the best unbiased linear predictor of log true income per capita based on log GDP per capita and log survey means. We are not the first to suggest or estimate a measure of true income as a linear combination of log GDP per capita and log survey means (Deaton 2001, Karshenas 2003, Chen and Ravallion 2010), but we contribute to this idea by proposing a data-driven method to estimate the optimal weights in this linear combination. We obtain that, if we do not include any controls, the best predictor of log true income per capita would place 85% of the weight on log GDP per capita and only 15% of the weight on log household survey means. Our more preferred specification includes fixed effects to allow surveys and national accounts to experience different kinds of biases across countries and years. In that specification, the best predictor of log true income places essentially 100% of the weight on log GDP per capita. We find that we should place nearly 100% of the weight on the national accounts measure when we look at different subsamples of the data, use national accounts consumption in place of GDP per capita, assign greater weight to poor countries with few surveys, use alternative measures of light intensity, and include a rich set of covariates that should capture any possible correlation between errors in nighttime lights and errors in national accounts GDP per capita and survey means.

This paper is closely related to the measurement error literature, including Adcock (1876), Griliches and Hausman (1982), Griliches (1986), Fuller (1987), Hausman (2001) and Chen, Hong and Nekipelov (2011), which have extensively studied the problem of uncovering structural parameters in systems of equations with measurement error. The most conceptually similar paper is Aruoba et al. (2014), which attempts to reconcile expenditure-based and income-based estimates of GDP growth in the U.S. by using unemployment as the independent referee variable, in the same way as we use lights. We believe that our paper is the first to use the econometric methods of accounting for measurement error in order to advance the debate between national accounts GDP per capita and survey means, and in particular, to construct an optimal combination of these two development indicators.

The rest of the paper is organized as follows. Section 2 describes the data that we use, including the lights measure. Section 3 describes our mathematical framework for computing optimal predictors of true income. Section 4 presents our results for the optimal linear combination of log national accounts GDP per capita and log household survey means, and states the central result that nearly all the weight should be on the national accounts. Section 5 presents the implications of our findings for estimates of global poverty. Section 6 describes a variant of our methodology that also includes lights directly as part of the proxy for true income. Section 7 presents an investigation of why the survey means appear to perform worse than the national accounts. Section 8 concludes.

2 Data

2.1 GDP

We use national accounts data from the World Bank (GDP per capita, PPP, constant 2005 international dollars).[3] The overwhelming majority of countries do not have missing data for this variable. National accounts data (from the World Bank or from the Penn World Tables) is widely used in cross-country studies of determinants of growth that follow Barro (1991).[4] We use data from the World Bank rather than from the Penn World Tables because of the known instability of the latter series (Ciccone and Jarocinski 2010; Johnson et al. 2013), and following the recommendation of Johnson et. al. (2013), who find that the World Bank series is constructed more consistently across time, although cross-country comparisons are more difficult to make.[5]

2.2 Survey Means

We use the dataset on mean survey income or consumption from household surveys collected by the World Bank (Povcalnet, http://iresearch.worldbank.org/PovcalNet/index.htm) and used by Chen and Ravallion (2001, 2004, 2010). All survey data is deflated to constant 2005 international dollars. This dataset mainly consists of surveys after 1990, although there are a few surveys present in the 1980s as well. Many of the survey parameters are heterogeneous (for instance, some surveys are income surveys and others are consumption surveys) but it appears that the heterogeneity is decreasing over time and is not particularly important for our results (allowing indicators for survey income concept does not affect our conclusions). On average, there are about 30-40 surveys each year since 1992, and there are 123 countries surveyed. Survey availability

[3] Before the current draft of this paper, but after the release of its working paper version, the ICP released the results of its 2011 price survey, and hence, new PPPs for the developing world. We continue to use 2005 PPPs because 1) the 2011 PPPs have not yet been incorporated into the World Bank's poverty estimates, and 2) for greater comparability with Chen and Ravallion (2010).

[4] These are, for example, Mankiw, Romer and Weil (1992), Barro (1999), Barro and Sala-i-Martin (2004), Sala-i-Martin, Doppelhoffer and Miller (2005), La Porta et al. (1999), Acemoglu et al. (2001, 2002, 2008), Spolaore and Wacziarg (2005), Ashraf and Galor (2013)

[5] An alternative could have been to use national accounts consumption per capita. We perform a robustness check using national accounts consumption in Section 4.3

is the primary constraint for our baseline sample from which to estimate the relative optimal weights of national accounts GDP per capita and survey means in the optimal proxy. Overall, we have 701 surveys in this sample, all of which match to national accounts and the lights data for the period 1992-2010. Chen and Ravallion (2010) present data on the fraction of population covered by surveys in each region in (or close to) each year.

Our sample contains observations from the developing world only: there are no World Bank surveys for OECD countries because OECD countries have virtually no population below the $1/day poverty line. We do not view this as a problem because we are mainly interested in measuring developing world living standards. Online Appendix Table II presents a list of all countries in the base sample, the number and date range of their surveys, and their income as measured by GDP, surveys and lights in the first and last year of their membership in the sample.

2.3 Nighttime Lights

Data on lights at night is collected by the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) satellite program and is maintained and processed by the National Oceanic and Atmospheric Administration (NOAA). Satellites orbit the Earth, sending images of every location between 65 degrees south latitude and 65 degrees north latitude at a resolution of 30 arcseconds (approximately 1 square km at the equator) at 20:30 to 22:00 local time.[6] The images are processed to remove cloud cover, snow and ephemeral lights (such as forest fires) to produce the final product available for download at

http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html

The nighttime lights data is available from 1992 to 2012, and we use the data up to 2010 because of the paucity of household surveys after that date that have already been made available for research.

Each pixel (1 square kilometer) in the luminosity data is assigned a digital number (DN) representing its luminosity. The DNs are integers that range from 0 to 63. We construct our lights proxy for aggregate income by summing up all the digital numbers across pixels

$$\text{Lights}_{j,t} = \sum_{i=1}^{63} i * (\#\text{ of pixels in country } j \text{ and year } t \text{ with DN} = i\ )$$

This formula has been used to aggregate the nighttime lights maps into lights-based indices for

[6] There are one or two satellites recording nighttime lights in each year, with an old satellite being retired and a new satellite being launched every few years. The satellites from which data is avaliable are as follows: the satellite F-10 (in orbit 1992-1994), F-12 (1994-1999), F-14 (1997-2003), F-15 (2000-2007), F-16 (2004-2009) and F-18 (2010-).

each country and year in nearly the entire literature on nighttime lights in economics, including Henderson, Storeygard and Weil (2012), Chen and Nordhaus (2011) and Michalopoulos and Papaioannou (2013, 2014). For years with multiple satellites available, we average the logarithms of our aggregate luminosity measure, following Henderson, Storeygard and Weil (2012).

It is very well established that lights are very strongly correlated with measures of economic activity, such as national accounts GDP, in levels and growth rates. Henderson, Storeygard and Weil (2012) provide these correlations, dramatic pictures of long-term differences in incomes (North vs. South Korea) as well as short-term fluctuations (the Asian financial crisis of 1997-8) reflected in lights. Michalopoulos and Papaioannou (2013, 2014) present evidence that nighttime light density in a sample of African villages is correlated with development indicators for these villages. We present a scatterplot of log national accounts GDP per capita and log household survey means against log nighttime lights per capita in Figure III of our paper. We see that both of these measures strongly increase with log nighttime lights per capita in a roughly linear fashion. The first two cells of Table I of our paper show that the regression coefficient of log nighttime lights on log GDP per capita is 1.160 (s.e. 0.063), and the R^2 of the regression is 0.72, suggesting that log GDP per capita explains roughly three-fourths of the variation in log nighttime lights per capita. The univariate relationship between nighttime lights and survey means is similar. In the remaining cells of the first two columns of Table I, we show that both log GDP per capita and log survey means continue to have a strong linear relationship with log nighttime lights per capita if year fixed effects or country fixed effects are included.

Our paper is closest in spirit to Henderson, Storeygard and Weil (2012) and Chen and Nordhaus (2011) in that it also considers the problem of optimally combining measures of economic activity. However, instead of using nighttime lights as a component of such a measure, we use it as an auxiliary variable to help uncover the correlation structure between the measures we do wish to use in our index. We also consider a different type of predictor for true income that do either Henderson, Storeygard and Weil (2012) or Chen and Nordhaus (2011), which allows us to make fewer assumptions on the data generating processes that we consider.

There are also well-known problems with the relation between nighttime lights and economic development, which we need to take into account. Pixels with DN equal to 0 or 63 are top- or bottom-censored. The light data also are affected by overglow and blooming: light tends to travel to pixels outside of those in which it originates, and light tends to be magnified over certain terrain types such as water and snow cover (Doll 2008). Given that we will compute national-level estimates of aggregate lights, it is unlikely that these sources of error will be large enough or sufficiently correlated with important variables that they will confound our analysis. Another problem may be that satellites age in space and are eventually retired.

Hence, they might give inconsistent readings from year to year, or new satellites may give fundamentally different readings from old ones. While some evidence of this problem exists, we will show in Section 3 that our calculations are supported by assumptions that allow nighttime lights to have all of the data problems described above, so long as nighttime lights are correlated with true income.

2.4 Other Data

We use a number of covariates to test the crucial maintained assumption of our paper; that nighttime lights are correlated with GDP per capita or with household survey means only through their joint correlation with true income (see the introduction and Section 3 below). These covariates are log electricity production per capita, log GDP per energy unit consumed, log oil rents, log shares of GDP in agriculture, manufacturing and services, log capital formation as percent of GDP, log export share, log import share, log general government expenditure share of GDP, log consumption share, the income share of the richest 10% and the income share of the poorest 50%, log percentage urban population, log percentage rural population, log total population, log area, and latitude and longitude of the capital city. The income share variables are from PovcalNet, while the area and capital city coordinates are from the CIA World Factbook. All other covariates are from the World Development Indicators. The covariates will be discussed at greater length in Section 4.

3 Mathematical Framework

Our goal is to find the best unbiased linear predictor of log true income per capita $y^*_{i,t}$, which is the log total per capita value added in country i and year t that we would compute if surveys or national statistical systems could record all income being earned. We will assume that this true income per capita is generated through some exogenous stochastic process that may not be stationary (if there is economic growth, for example). We cannot observe $y^*_{i,t}$ directly. Instead, we can observe data on log light intensity per capita ($y^L_{i,t}$), log measured GDP per capita ($y^G_{i,t}$), and log survey mean income ($y^S_{i,t}$) for a sample of countries i and years t. These data are related to log true income per capita according to the following system of equations (partialling out constants and other covariates):

$$y^L_{i,t} = \beta^L y^*_{i,t} + \varepsilon^L_{i,t} \tag{1}$$

$$y^G_{i,t} = \beta^G y^*_{i,t} + \varepsilon^G_{i,t} \tag{2}$$

$$y_{i,t}^{S}=\beta^{S}y_{i,t}^{*}+\varepsilon_{i,t}^{S} \tag{3}$$

In other words, each of the measured variables is a linear function of log true income per capita, perturbed by some error.[7] This framework is very similar to the one used by Henderson, Storeygard and Weil (2012) and Chen and Nordhaus (2011), except that both of these papers assume $\beta^{G}=1$, which means that log GDP per capita is an unbiased proxy for log true income per capita. Instead, we do not assume that any of our measured proxies are unbiased, and allow them to deviate from log true income per capita along a linear trend. This implies that, for instance, if β^{S} is less than β^{G}, then household surveys, on average, capture a smaller fraction of economic growth than is recorded in the national accounts, causing a persistent divergence between survey means and national accounts GDP per capita over time. While this property would appear to be surprising, it is actually consistent with the data. In fact, the ratio of household survey means to national accounts GDP per capita has been declining over the past few decades (Deaton 2005), and we allow for the coefficients β^{G} and β^{S} to differ precisely in order to capture this feature of the data. Within our framework, we can test and reject the null hypothesis that the coefficients in the GDP and the surveys equation are equal ($\beta^{S}/\beta^{G}=1$), a necessary condition for unbiasedness of both measures simultaneously, in favor of the hypothesis $\beta^{S}/\beta^{G}<1$, which is the necessary and sufficient condition to generate the empirical fact that the ratio between survey means and national accounts GDP per capita declines with economic growth over time.[8]

We assume that the error terms in all three processes are mean independent of true income. That is,

$$E\left(\varepsilon_{i,t}^{L}|y_{i,t}^{*}\right)=E\left(\varepsilon_{i,t}^{G}|y_{i,t}^{*}\right)=E\left(\varepsilon_{i,t}^{S}|y_{i,t}^{*}\right)=0 \tag{A1}$$

The critical assumption of this paper is that the error term in the lights equation (1), $\varepsilon_{i,t}^{L}$, is uncorrelated with the error terms from the GDP and surveys equations (2) and (3), $\varepsilon_{i,t}^{G}$ and $\varepsilon_{i,t}^{S}$ conditional on

[7] In fact, we can substantially relax the functional form specification in equation 1 to read

$$y_{i,t}^{L}=f_{i,t}\left(y_{i,t}^{*}\right)+\varepsilon_{i,t}^{L}$$

so long as

$$cov\left(y_{i,t}^{*},f_{i,t}\left(y_{i,t}^{*}\right)\right)\neq 0$$

(The analogous assumption in our framework is $\beta^{L}\neq 0$. We test and confirm both of these assumptions in Section 4).

This is a much more general framework that allows for errors in the lights data such as nonlinearity, top- and bottom-coding, differences in the lights-to-income relationship as satellites age and are replaced, and differences in the lights-to-income relationship across countries because of cultural attitudes to nighttime light and light pollution.

[8] Of course, if β^{S} is always less than β^{G}, in the limit, the ratio of household survey income to national accounts GDP will go to zero and household surveys will be trivial. This will likely cause the statistical agencies to change the way that they conduct household surveys, which will result in a different β^{S} that is closer to β^{G}. In that case, the parameter β^{S} would be time-varying and our model as specified in equations 1-3 would not be correct. However, we can accommodate such a process by reestimating the model separately for different time periods. While we do not find evidence of changes in the ratio β^{S}/β^{G} over time in our sample, this does not preclude such changes in the future.

true income:

$$E\left(\varepsilon_{i,t}^{G}\varepsilon_{i,t}^{L}|y_{i,t}^{*}\right)=E\left(\varepsilon_{i,t}^{S}\varepsilon_{i,t}^{L}|y_{i,t}^{*}\right)=0 \tag{A2}$$

Assumption A2 is the key reason for the use of the lights data. This assumption has also been made in Henderson, Storeygard and Weil (2012) and Chen and Nordhaus (2011). This is a plausible assumption because the data generating processes of the lights data and of GDP (or surveys) are largely disjoint; lights data is collected by satellites without respect for borders, institutional structures, or people's desire to respond to surveys, whereas GDP and survey data are obtained primarily or largely by asking people, who may be unwilling or unable to respond accurately. We will extensively investigate the robustness of our results to this assumption.

We are interested in finding the best unbiased linear predictor of log true income per capita ($y_{i,t}^{*}$) in terms of $y_{i,t}^{G}$ and $y_{i,t}^{S}$,

$$z_{i,t}=\gamma_{G}y_{i,t}^{G}+\gamma_{S}y_{i,t}^{S} \tag{4}$$

Hence, we want to compute the vector of weights (γ_{G},γ_{S}), which minimizes the mean squared error

$$(\gamma_{G}^{*},\gamma_{S}^{*})=\arg\min_{\gamma_{G},\gamma_{S}}E\left(\left(y_{i,t}^{*}-\gamma_{G}y_{i,t}^{G}-\gamma_{S}y_{i,t}^{S}\right)^{2}\right) \tag{5}$$

subject to the constraint that the proxy be unbiased, that is:

$$E\left(\gamma_{G}^{*}y_{i,t}^{G}+\gamma_{S}^{*}y_{i,t}^{S}|y_{i,t}^{*}\right)=y_{i,t}^{*} \tag{6}$$

This constraint can be reformulated as

$$\gamma_{G}^{*}\beta^{G}+\gamma_{S}^{*}\beta^{S}=1 \tag{7}$$

Now, plugging equations (2) and (3) into the value function equation (5), we obtain

$$\begin{aligned}&E\left(\left(y_{i,t}^{*}-\gamma_{G}y_{i,t}^{G}-\gamma_{S}y_{i,t}^{S}\right)^{2}\right)\\=&\left(1-\gamma_{G}^{*}\beta^{G}-\gamma_{S}^{*}\beta^{S}\right)E\left(\left(y_{i,t}^{*}\right)^{2}\right)+\gamma_{G}^{2}var\left(\varepsilon_{i,t}^{G}\right)+2\gamma_{G}\gamma_{S}cov\left(\varepsilon_{i,t}^{G},\varepsilon_{i,t}^{S}\right)+\gamma_{S}^{2}var\left(\varepsilon_{i,t}^{S}\right)\end{aligned} \tag{8}$$

and applying the constraint equation (7) eliminates the first term.

Therefore, our constrained optimization problem becomes

$$\left(\gamma_G^*, \gamma_S^*\right) = \arg\min_{\gamma_G, \gamma_S} \left\{\gamma_G^2 var\left(\varepsilon_{i,t}^G\right) + 2\gamma_G\gamma_S cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right) + \gamma_S^2 var\left(\varepsilon_{i,t}^S\right)\right\} \tag{9}$$

subject to

$$\gamma_G^* \beta^G + \gamma_S^* \beta^S = 1 \tag{10}$$

Solving this problem with traditional constrained optimization techniques, we obtain the relation

$$\frac{\gamma_G^*}{\gamma_G^* + \gamma_S^*} = \frac{\beta^G var\left(\varepsilon_{i,t}^S\right) - \beta^S cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right)}{\beta^G var\left(\varepsilon_{i,t}^S\right) + \beta^S var\left(\varepsilon_{i,t}^G\right) - \left(\beta^G + \beta^S\right) cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right)} \tag{11}$$

The expression $\frac{\gamma_G^*}{\gamma_G^* + \gamma_S^*}$ is the weight that should be given to log GDP per capita relative to the total weight that should be placed on log GDP per capita and log household survey means in the best unbiased linear predictor of log true income per capita in equation (4). However, the moments on the right hand-side of equation (11) are unknown. We now proceed to show that we can use the lights data to recover this function of unknown moments. To do so, we compute the population regression

$$y_{i,t}^L = b^0 + b^G \; y_{i,t}^G + b^S \; y_{i,t}^S \tag{12}$$

which is the regression of log lights per capita on log GDP per capita and log survey means. We then show that the population regression coefficients that we obtain satisfy

$$\frac{b^G}{b^G + b^S} = \frac{\gamma_G^*}{\gamma_G^* + \gamma_S^*} \tag{13}$$

In other words, the regression coefficient on log GDP per capita, divided by the sum of the regression coefficients on log GDP per capita and log survey means yields us the relative weight of log GDP per capita in the best unbiased linear predictor (4), which is exactly what we want. So all that we have to do is to run a linear regression.

To prove the equality (13), we note that the formula for the lights regression coefficient on log GDP per capita, b^G, is as follows:

$$b^G = \frac{var\left(y_{i,t}^S\right) cov\left(y_{i,t}^G, y_{i,t}^L\right) - cov\left(y_{i,t}^G, y_{i,t}^S\right) cov\left(y_{i,t}^S, y_{i,t}^L\right)}{var\left(y_{i,t}^G\right) var\left(y_{i,t}^S\right) - \left(cov\left(y_{i,t}^G, y_{i,t}^S\right)\right)^2} \tag{14}$$

The formula for the lights regression coefficient on log survey means, b^S, is analogous, switching around

$y^S_{i,t}$ and $y^G_{i,t}$ in equation (14). These formulas depend on the variances and covariances of $y^L_{i,t}$, $y^G_{i,t}$ and $y^S_{i,t}$. Under Assumptions A1 and A2, as well as equations (1)-(3), these variances and covariances can be expressed as follows:[9]

$$var\left(y^G_{i,t}\right) = \left(\beta^G\right)^2 var\left(y^*_{i,t}\right) + var\left(\varepsilon^G_{i,t}\right) \tag{15}$$

$$var\left(y^S_{i,t}\right) = \left(\beta^S\right)^2 var\left(y^*_{i,t}\right) + var\left(\varepsilon^S_{i,t}\right) \tag{16}$$

$$cov\left(y^G_{i,t}, y^L_{i,t}\right) = \beta^G\beta^L var\left(y^*_{i,t}\right) \tag{17}$$

$$cov\left(y^S_{i,t}, y^L_{i,t}\right) = \beta^S\beta^L var\left(y^*_{i,t}\right) \tag{18}$$

$$cov\left(y^G_{i,t}, y^S_{i,t}\right) = \beta^G\beta^S var\left(y^*_{i,t}\right) + cov\left(\varepsilon^G_{i,t}, \varepsilon^S_{i,t}\right) \tag{19}$$

Substituting equations (15)-(19) into equation (14) leads to the formula[10]

$$b^G = \frac{\beta^L var\left(y^*_{i,t}\right)\left[\beta^G var\left(\varepsilon^S_{i,t}\right) - \beta^S cov\left(\varepsilon^G_{i,t}, \varepsilon^S_{i,t}\right)\right]}{var\left(y^*_{i,t}\right)\left[var\left(\beta^G\varepsilon^S_{i,t} - \beta^S\varepsilon^G_{i,t}\right)\right] + var\left(\varepsilon^S_{i,t}\right) var\left(\varepsilon^G_{i,t}\right) - cov\left(\varepsilon^G_{i,t}, \varepsilon^S_{i,t}\right)} \tag{20}$$

The result for the coefficient b^S is analogous, switching around β^G and β^S in the numerator and replacing $var\left(\varepsilon^S_{i,t}\right)$ with $var\left(\varepsilon^G_{i,t}\right)$ in the numerator.

If we divide equation (20) by $b^G + b^S$, we obtain[11]

[9] In the more general model in which we assume a general functional form for the lights-true income relation (see footnote 8), equations (17) and (18) replace the term $\beta^L var\left(y^*_{i,t}\right)$ with the term $cov\left(y^*_{i,t}, f_{i,t}\left(y^*_{i,t}\right)\right)$.

[10] In the more general model in footnote 8, equation (20) should replace the term $\beta^L var\left(y^*_{i,t}\right)$ with the term $cov\left(y^*_{i,t}, f_{i,t}\left(y^*_{i,t}\right)\right)$. The rest of the logic proceeds exactly as in the text.

[11] We may divide by $b^G + b^S$ in equation (20) so long as a) $\beta^L \neq 0$ (or $cov\left(y^*_{i,t}, f_{i,t}\left(y^*_{i,t}\right)\right) \neq 0$ in the more general model), and b) the inequality

$$\beta^G var\left(\varepsilon^S_{i,t}\right) + \beta^S var\left(\varepsilon^G_{i,t}\right) - \left(\beta^G + \beta^S\right) cov\left(\varepsilon^G_{i,t}, \varepsilon^S_{i,t}\right) \neq 0$$

holds. Condition a) holds so long as lights have some relation to true income, and can be tested because its failure implies that $b^G = b^S = 0$. Condition b) holds for generic parameter values of the model as long as i) either GDP per capita or survey means have some relation to true income, that is $\left(\beta^G, \beta^S\right) \neq (0,0)$ ii) GDP per capita and survey means are not identical random variables. We can easily test that $b^G + b^S \neq 0$ when we run the regression (12).

$$
\begin{aligned}
\frac{b^G}{b^G+b^S} &= \frac{\beta^G var\left(\varepsilon_{i,t}^S\right)-\beta^S cov\left(\varepsilon_{i,t}^G,\varepsilon_{i,t}^S\right)}{\beta^G var\left(\varepsilon_{i,t}^S\right)+\beta^S var\left(\varepsilon_{i,t}^G\right)-\left(\beta^G+\beta^S\right) cov\left(\varepsilon_{i,t}^G,\varepsilon_{i,t}^S\right)} \\
&= \frac{\gamma_G^*}{\gamma_G^*+\gamma_S^*}
\end{aligned}
$$

Therefore, we can estimate the weight that should be placed on log GDP per capita relative to the total weight on the two proxies for log true income per capita. We cannot estimate γ_G^* and γ_S^* individually, or the sum $\gamma_G^*+\gamma_S^*$, but we can estimate their ratios. Intuitively, we can estimate what fraction of the weight that log GDP per capita should receive, but we cannot estimate how large this "weight" is.

The core of our analysis in Section 5 will be running regressions similar to equation (12) presenting estimates of the optimal weight on GDP per capita $\frac{b^G}{b^G+b^S}$ and the optimal weight on the survey means $\frac{b^S}{b^G+b^S}$ when the elementary specifications in equations (1), (2) and (3) – and hence, regression equation (12) – are augmented by covariates, when they are estimated on different samples, or when when they are estimated using different lights measures.

As mentioned above, the critical assumption of our analysis is Assumption A2: the orthognality of the error in lights to the errors in GDP and surveys. Therefore, it is important to ask about how violations of this assumption might affect our analysis. One possible scenario may be if developing countries use government estimates of electricity production, which is obviously correlated with nighttime lights, to calculate GDP. Another possibility may be if the outputs of industries such as manufacturing, or of activities such as investment (construction) are more light-intensive per unit of GDP produced than other activites, and also if they are more easily measured with national accounts than with household surveys.[12] However, errors in lights may also be correlated with errors in household surveys. For example, both household surveys and nighttime lights may systematically fail to capture top incomes (the former because of misreporting and the latter if highly concentrated incomes do not generate much light-producing activity), while national accounts may capture production that eventually constitutes the incomes of the very rich. From equation (14), we can easily see that the estimate of the optimal relative weight on national accounts, $\frac{b^G}{b^G+b^S}$, is increasing in the ratio of the covariances $\frac{cov\left(y_{i,t}^G,y_{i,t}^L\right)}{cov\left(y_{i,t}^S,y_{i,t}^L\right)}$. Therefore, our estimates will put too much weight on GDP per capita whenever

$$
\frac{\beta^S}{\beta^G} cov\left(\varepsilon_i^G,\varepsilon_i^L\right) \geq cov\left(\varepsilon_i^S,\varepsilon_i^L\right) \tag{21}
$$

and they will put too much weight on household surveys if this relation fails. For example, if national statistical offices estimate GDP using electricity production, we should expect $cov\left(\varepsilon_i^G,\varepsilon_i^L\right)>0$, and if survey

[12] We thank Angus Deaton for bringing these particular examples to our attention.

errors are uncorrelated with lights ($cov\left(\varepsilon_i^S, \varepsilon_i^L\right) = 0$), the inequality (21) holds and our estimates place too much weight on GDP. On the other hand, if both household surveys and nighttime lights understate top incomes while GDP captures them, $cov\left(\varepsilon_i^S, \varepsilon_i^L\right) > 0$ and $cov\left(\varepsilon_i^G, \varepsilon_i^L\right) = 0$, so inequality (21) fails and our estimates place too much weight on household surveys.

We alleviate these concerns in Section 4 by noting that violations of Assumption A2 would be mediated by variables that we can measure, such as electricity consumption per capita, sectoral shares of national income, urbanization, inequality and others, all of which we can flexibly include as control variables in our equations (1)-(3). We find that upon accounting for these controls, our analysis is very little changed. We believe that these controls affect our analysis so marginally because we are already implicitly accounting for them by allowing a linear bias in the relations between the measured variables and log true income when we do not restrict the coefficients β^L, β^G and β^S to equal unity.

4 Results for Optimal Weights

In this section, we will use GDP per capita from the national accounts, household survey means, satellite data on nighttime lights as well as conditioning variables to check robustness to possible violations of our assumptions in order to estimate the ratio of the weight on survey means to that of national accounts GDP per capita in the optimal proxy.

4.1 Regressions of Nighttime Lights on GDP per Capita and Survey Means

It is important to verify explicitly that there indeed exist relationships between nighttime lights, national accounts GDP per capita and true income. As mentioned in Section 2, Table I presents univariate regressions of log nighttime lights per capita on log GDP per capita and on log survey means, as well as bivariate regressions of nighttime lights on both GDP per capita and survey means, for our base sample of 701 country-years in the developing world with survey information. It can easily be shown that under Assumptions A1 and A2 in Section 3, the coefficients in the univariate regressions are proportional to the expressions $\beta^L\beta^G$ and $\beta^L\beta^S$ respectively, so they are positive and significant if and only if both lights and GDP (or lights and surveys) have statistically significant relationships of identical sign with true income per capita. Hence, the univariate regressions are a basic check of our assumption that β^L, β^G and β^S can all be taken to be greater than zero.[13] The first two cells of Table I provide the regression coefficients of log lights per capita on log GDP per capita and of log lights per capita on log survey means, respectively. We can see that both are

[13] Under the more general model, these coefficients are proportional to $\beta^G cov\left(y_{i,t}^*, f_{i,t}\left(y_{i,t}^*\right)\right)$ and $\beta^S cov\left(y_{i,t}^*, f_{i,t}\left(y_{i,t}^*\right)\right)$ respectively.

statistically significant and large. Hence, our assumption that nighttime lights per capita, national accounts GDP per capita and survey means are strongly associated with true income per capita is not falsified.

The procedures of the national statistical agencies that compile the national accounts and administer the surveys may vary systematically across countries, and may also vary systematically over time because of improvements in national accounting. Since the quality of satellites that record nighttime lights changes from year to year, and since different countries may have different light-generating processes, there is a worry that the country- and year-specific effects in log national accounts GDP per capita and log household survey means may be correlated with country- and year-specific effects in log nighttime lights per capita. To address this problem, we estimate our regression adding year fixed effects, country fixed effects, and both, in the remaining three panels of Table I. We see in column 1 that the correlation between log nighttime lights per capita and log GDP per capita remains strong and statistically significant regardless of the fixed effects included. This finding justifies ex post our readings of Figures I and II, the pictures of India and southern Africa, in which we interpreted changes in lights over time as indicative of economic growth. However, in column 2, we see that the association between log nighttime lights per capita and log household survey means weakens once we include country fixed effects and disappears (shrinks to one-tenth of its value and loses significance) when we include both country and year fixed effects. Hence, while the levels of log survey means are correlated with the levels of lights per capita as strongly as are the levels of log GDP per capita, the growth rates of survey means have a much weaker correlation with the growth rates of nighttime lights per capita, while the growth rates of GDP per capita are tightly correlated with them. This observation shows that Figures I and II are not coincidences, but illustrate a broad statistical pattern: the growth rate of lights is much more similar to the growth rate of GDP per capita than it is to the growth rate of household survey means.

We can see the strong, linear relationships between log lights per capita and log national accounts GDP per capita, as well as between log lights per capita and log household survey means in Figure III, which plots the data underlying the correlations in the first two cells of Table I. We immediately observe that the regression coefficients in these cells are not driven by outliers or nonlinearities, but rather that the linear relationships are the major feature of the data in question. We also observe that the slopes of the scatterplots are unmistakeably different, with the distance between the trendline based on log GDP per capita and the trendline based on log survey means much larger for richer countries than for poorer ones. Since Figure III plots log nighttime lights on the x-axis, and log GDP per capita and log survey means on the y-axis, these slopes should be the same if and only if $\beta^G = \beta^S$. Hence, Figure III graphically shows that $\beta^G > \beta^S$, which justifies our modeling of log GDP per capita and log survey means as increasing at different rates with log true income.

We can test this hypothesis more formally. Equations (17) and (18) readily show that testing $\beta^S/\beta^G = 1$ is equivalent to testing the hypothesis

$$\frac{cov\left(y^S_{i,t}, y^L_{i,t}\right)}{cov\left(y^G_{i,t}, y^L_{i,t}\right)} = 1$$

The expression on the left hand-side is just the slope coefficient in an instrumental variables regression of log survey mean income on log GDP per capita, instrumenting the regressor with log nighttime lights per capita. We report the estimates and standard errors of this coefficient in column 3 of Table I for the different fixed effects specifications that we consider. We see that we can always reject the null hypothesis $\beta^S/\beta^G = 1$ against the alternative that $\beta^S/\beta^G < 1$, with the upper bound on the 95% confidence interval of the coefficient of instrumented log GDP per capita being 0.9. Hence, allowing $\beta^S < \beta^G$ is essential to capturing an important feature of our data.

Column 4 of Table I presents a preview of the main result of this paper by estimating the regression equation (12). As discussed in Section 3, the coefficients in a bivariate regression of log lights per capita on log GDP per capita and log survey means are proportional to the weights γ^*_G and γ^*_S in the best (minimum-variance linear unbiased) lights-based proxy for log true income per capita. This is intuitive because equation (1) and assumption A2 imply that log lights per capita are just log true income per capita multiplied by a coefficient and perturbed by some noise that is uncorrelated with either of the errors in national accounts GDP per capita or survey means. The first cell shows that once both log GDP per capita and log survey means are included in the no fixed effects regression, the coefficient on log GDP remains close to the univariate regression – it is 1.020, and significant at 1% – while the coefficient on survey means collapses by nearly a factor of ten to an insignificant 0.184. Hence, in the bivariate regression, log GDP per capita wins the horse race easily. We illustrate this fundamental result graphically in Figure IV, which plots the partial correlations between log lights per capita and log national accounts GDP per capita, as well as between log lights per capita and log household survey means. Once again, the strong partial correlation between log lights per capita and log national accounts GDP per capita is not driven by outliers or nonlinearities. Similarly, the absence of a partial correlation between log lights per capita and log household survey means cannot be explained by a few observations, but instead is a major feature of the data. Including year, country, and all fixed effects only reinforces our conclusion. Country fixed effects turn the coefficient on log survey means negative (and insignificantly different from zero), but do not affect the significance of the coefficient on log GDP per capita, though they decrease its magnitude somewhat.

An important concern is that statistical agencies use electricity to compute GDP (Deaton, personal communication), which can bias the data in favor of producing our multivariate regression results, as we discussed in Section 3. A straightforward way of alleviating this concern is to include log electricity production per capita as a control in regression (12).[14] Column 5 of Table I presents the resulting regression estimates for all four of the specifications we consider. We see that the coefficient on log GDP per capita remains statistically significant and large in all specifications, but declines in magnitude to about half of what it was in column 4. However, the coefficient on log household survey means also declines in magnitude nearly to zero, and is not statistically significant. Therefore, the qualitative conclusions of the simple multivariate regression analysis in column 4 remain unchanged, even though it is true that electricity production explains some (but not all) of the correlation between log GDP per capita and log nighttime lights per capita. Additionally, log GDP per capita and log electricity production per capita are positively correlated, which means that the coefficient on log GDP per capita is measured less precisely than it was in the regression without log electricity per capita. In particular, in the country fixed effects specification, the t-statistic on log GDP per capita declines from over 4 to just a little over 2.

4.2 Estimates of Relative Weights

Let us define the unobservable relative weight of log GDP per capita in the best unbiased linear proxy for true income as

$$\omega_G^* = \frac{\gamma_G^*}{\gamma_G^* + \gamma_S^*} \tag{22}$$

in the notation of Section 3, and let us define the ratio of the observable coefficient on log GDP per capita to the sum of the coefficients from regression (12) as

$$\hat{\omega}_G = \frac{b^G}{b^G + b^S} \tag{23}$$

The relative weight of log survey means, ω_S^*, is just given by $1 - \omega_G^*$, and the corresponding observable statistic is $\hat{\omega}_S = 1 - \hat{\omega}_G$.

The central result of Section 3 is that $\hat{\omega}_G$ is a consistent estimator of ω_G^*. Research using exclusively national accounts implicitly assumes that $\gamma_G^* = 1$, and hence that $\omega_G^* = 1$ (the literature following Barro 1991). Research that exclusively uses survey means implicitly assumes that $\gamma_G^* = 0$ and hence that $\omega_G^* = 0$ (Chen and Ravallion 2004, Milanovic 2005). Chen and Ravallion (2010) consider a mixed method in which they measure income per capita by the geometric mean of the survey mean consumption and the fitted value of survey mean consumption from a regression of log consumption on a constant and on log consumption in the national accounts. They report that the coefficient on log consumption from the national accounts in

[14] In terms of the model in Section 3, this is equivalent to including log electricity consumption per capita in the measurement equations (1)-(3), the conditioning variables in Assumptions A1 and A2, and the formula (4) for the lights-based predictor $z_{i,t}$.

such a regression tends to be between 0.6 and 0.85, so we can consider the Chen-Ravallion proxy to be given by

$$z_i^{CR} = \frac{1}{2} y_i^S + \frac{1}{2} \rho y_i^G$$

where $\rho \in (0.6, 0.85)$. Hence, the Chen-Ravallion (2010) approach assumes that $\omega_S^* > \omega_G^*$, and hence that $\omega_G^* < 0.5$.[15]

The central part of our paper will be to use the statistic $\hat{\omega}_G$ to see which of these null hypotheses are correct. Table II presents the statistics $\hat{\omega}_G$ and $\hat{\omega}_S$ for a variety of specifications. It is important to recognize that while $\hat{\omega}_G$ converges in probability to ω_G^* as the sample size goes to infinity so long as $\gamma_G^* + \gamma_S^* > 0$[16], the performance of this estimate in finite samples may be nonstandard. In particular, $\hat{\omega}_G$ is a quotient of normally distributed regression coefficients, and therefore may be poorly estimated if the variances of these coefficients are large. In particular, $var\left(\varepsilon_{i,t}^L\right)$, the variance of the error term in the relation between lights and true income, is part of the error term of the regression equation (12). If lights are a noisy proxy for true income (which is the conclusion of HSW 2012 and CN 2010), then the variance of the sum $b^G + b^S$ can be expected to be large, which would make inference based on an appeal to asymptotic theory dubious. Therefore, instead of reporting asymptotic standard errors for $\hat{\omega}_G$, we provide 95% confidence intervals for this statistic based on 200 block-bootstrapped samples in which the bootstrap block is the country. As we will see, these confidence intervals need not be symmetric around the estimated value of $\hat{\omega}_G$, but appear to be relatively tight, as would be expected in the reasonably rich sample that we have (700 observations in over 120 clusters). Hence, we alleviate the concern that our inference may be insufficiently conservative because $\hat{\omega}_G$ has a nonstandard distribution.

Our simplest specification (Row 1 and Column 1 of Table II) estimates the regression equation (12) without any controls or fixed effects. The first part of the cell presents $\hat{\omega}_G$ and its 95% confidence interval. It suggests that the relative weight of log GDP per capita in the best unbiased proxy is 0.84, and that with 95% confidence, it is between 0.64 and 1.05. Note that the number 0.5 is not inside this interval, so we easily reject the null hypothesis that $\omega_G^* = 0.5$, or surveys get the same weight as national accounts, and a fortiriori, we reject that $\omega_G^* = 0$, or that all the weight should be placed on the surveys. We also *fail* to reject the null hypothesis that $\omega_G^* = 1$, or surveys get zero weight in the optimal proxy, while national accounts get all the weight. The second part of the first cell contains the estimate on the relative weight of log survey means $\hat{\omega}_S$, which is just $1 - \hat{\omega}_G$. We see that $\hat{\omega}_S = 0.15$, and that its confidence interval contains 0, but not 0.5 or 1, thus leading to exactly the same inferences as we had earlier.

[15] To be more precise, given that Chen and Ravallion (2010) note that $\rho \leq 0.85$, the relevant hypothesis is actually $\omega_G^* < 0.46$. However, we typically reject the stronger null that $\omega_G^* < 0.5$.

[16] A sufficient condition for which is $\beta^L > 0$ and at least one of β^G and β^S is greater than 0, which we have tested empirically through the univariate regressions in Table I

The rest of column 1 of Table II shows estimates of ω_G^* and ω_S^* with various types of fixed effects included into our specification. For column 1, the baseline specification (log lights per capita measure and no controls), the results strengthen slightly in favor of log GDP per capita as we include more fixed effects. In particular, when we include country fixed effects, or country-year fixed effects, the estimate of ω_G^* becomes 1.09, with a 95% confidence interval of $(0.77, 1.3)$. Hence, if we include country fixed effects, our best point estimate for how to construct the optimal proxy for true income is to exclude household surveys completely. More generally, whatever variation we use to identify the weight on the optimal proxy – cross-country income distribution variation, or variation in growth rates between and within countries, or even business cycle variation between and within countries – the estimates for the relative weights that we obtain are largely the same. In particular, differences in the assumptions of statistical agencies in the construction of GDP or in the implementation of the surveys, which should vary by country but much less so by year, cannot be driving our results.

4.3 Robustness Checks

In columns 2-4 of Table II we augment our baseline regression equation (12) with various controls. The reason why we may need controls in our specification is the possible failure of Assumption A2: the concern that national accounts GDP per capita and survey means may be correlated with lights for other reasons than their joint correlation with true income. For example, as suggested in Section 4.1, some developing countries may use estimates of electricity production as the basis for their estimates of GDP, and electric lighting is a major part of observed nighttime lights. In Column 2, we include log electricity production per capita (from the WDI) as a control. We observe that our point estimates of the optimal relative weight on GDP per capita, ω_G^*, remain the same or become larger, although their confidence interval expands so that we cannot reject $\omega_G^* = 0.5$ in the two specifications with country fixed effects.[17] To understand what is going on, we recall that in the fixed effects regression that generated the coefficients b^G and b^S, in Column 5 of Table I, adding log electricity per capita as a control decreased the t-statistics on these coefficients because of multicollinearity between log GDP per capita, log survey means and log electricity consumption per capita. The higher standard errors on the coefficients b^G and b^S contribute to even wider confidence intervals for ω_G^* because it is a quotient of normally distributed variables, and therefore, quite sensitive if its denominator shrinks.

Column 3 controls for other potential confounders of the relationship between nighttime lights and GDP alongside electricity production per capita. Specifically, these confounders are:

[17] We can reject this null hypothesis in the no fixed effects specification. For the specification with year fixed effects, we marginally reject the null hypothesis $\omega_G^* = 0.46$, which is the effective null hypothesis of the mixed method in Chen and Ravallion (2010).

- Oil rents as percent of GDP (because oil wells generate large amounts of light)
- GDP per energy unit consumed (because this will obviously change the relation between true income and lights)
- Shares of GDP in agriculture, manufacturing and services (because manufacturing may be more light-intensive than the other two sectors).
- General government expenditure share of GDP (because government goods, such as military technology, may be more light-intensive)
- Shares of GDP in exports and imports (because they are measured particularly well in national accounts and may generate large amounts of light through ports and warehouses).
- Income shares of the richest 10% and the poorest 50% (because light may be a necessity, and the consumption of the rich may generate less light; alternatively, the consumption of the poor may generate little light if they aren't electrified).
- Capital formation as percent of GDP (because capital may be particularly light-intensive)
- Consumption share of GDP (because consumption might not be very light-intensive)
- Population (because higher population density almost always entails more light)
- Fractions of the population rural and urban (because urban settings generate more light per capita, through infrastructure)
- Area (both total and arable, because small areas can be associated with high population densities)
- Latitude and longitude of the capital city (because geographic location affects climate, and thus measurement errors in lights).

Lastly, in column 4, we include all of the above controls (as well as log electricity per capita) together with their squares in order to capture any potential nonlinearities in their relationship with nighttime lights. Once again, the inferences are unchanged, and we see that our estimates of ω_G^*, if anything, are larger, suggesting an even greater role for national accounts in the best proxy for true income. For example, in the no fixed effects specification (row 1), the estimate of ω_G^* when all the controls and their squares are included (column 4) is 0.99, which is close to the estimate of ω_G^* when country fixed effects are included. This observation might suggest us to place greater confidence in the country fixed effects specification, because it appears likely that the country fixed effects proxy for omitted variables in the relationships between true income, national accounts GDP per capita, survey means and lights, and controlling for these omitted variables makes it more likely that Assumptions A1 and A2 hold. (When we include controls in the fixed effects specification, the point estimate of ω_G^* rises trivially from 1.09 to 1.15, so it appears that country fixed effects and the rich set of covariates are largely proxying for the same omitted variables). We are further reassured that even if Assumptions A1 and A2 appear to require additional controls or fixed effects to hold, making these assumptions in the more parsimonious model of Section 3 biases our estimates towards finding a greater weight for log household survey means in the optimal proxy for true income, and away from the main conclusion of the paper.[18]

Our measurement equation (1) relates lights to true income, but does not explain how to measure lights. In Section 2, we describe the way in which we transform the light pixel maps produced by the NOAA into numerical indices of lights per capita for countries in different years. We believe that measuring aggregate lights per capita is the best counterpart to GDP per capita or survey means. However, alternative ways of obtaining a lights proxy for economic activity are possible, and have been used in the literature (Henderson, Storeygard and Weil 2012, Michalopoulos and Papaioannou 2013, Pinkovskiy 2014). Columns 5-7 of Table II experiment with various alternative ways of constructing light indices. Column 5 presents results using light density (aggregate lights per area) rather than lights per capita as in our baseline[19] and column 6 presents results using a modified aggregate radiance measure in which the exponent on the digital number (unity in the baseline measure) is calibrated so as to match as closely as possible the average income of the states of Mexico, obtained from the Luxembourg Income Study.[20] We see that neither measure produces results radically different from the baseline. Column 7 uses disaggregated population data from the Gridded Population of the World to compute the fraction of each country's population living in areas with observed lights. As this disaggregated population data is available only at 5-year intervals, our sample size shrinks dramatically (to 160 observations), which causes confidence intervals to widen when we include country and year fixed effects. However, all the estimates of the relative weight of national accounts, ω_G^*, are close to the baseline estimates (between 0.8 and 1.12), and our inferences are almost completely unchanged.

After checking alternative measures for lights, we proceed to checking alternative national accounts and survey measures. There is controversy over which concept of economic activity best corresponds to income. Anand and Segal (2008) argue that GDP per capita is conceptually different from disposable income, and that a better national accounts measure is household final consumption expenditure (HFCA). Moreover,

[18] Of course, our analysis is merely consistent with this interpretation of the omitted variables bias, and other interpretations are possible.

[19] We use this measure of lights for completeness, following papers in the nighttime lights literature. In specifications without fixed effects, measuring lights per area can lead to understating the development of a country because it has a large area.

[20] We allow the calibrated exponent to differ across years, but in no year is it smaller than 5/2, and in some years it is as large as 9. Therefore, it is likely that the specification that is prevalent in the literature (setting the exponent equal to unity) is incorrect.

household surveys are very heterogeneous, with some measuring income and others measuring consumption, and it may be more appopriate to compare survey consumption measures with national accounts consumption measures, such as HFCA, rather than with GDP. Columns 8-11 of Table II check the robustness of our results to using national accounts consumption in place of GDP, and to accounting for the heterogeneity in survey methodology. Column 8 replicates the baseline estimates replacing log national accounts GDP per capita with log national accounts consumption per capita.[21] We obtain virtually the same results: the weight on log national accounts consumption per capita varies between 0.82 and 1.03 depending on whether fixed effects are included or excluded, and we can always reject the null that national accounts should get zero weight (and fail to reject the null that they should get all the weight). Columns 9 and 10 estimate the model in Section 3 separately for surveys that measure household income (column 9, in which case we combine them with GDP per capita), and for surveys that measure consumption (column 10, in which case, we combine them with national accounts consumption). Most surveys (438 out of 701) actually measure household consumption rather than income, so the sample of income surveys is small ($N = 263$), which leads to wide confidence intervals in that specification. Once again, we estimate that national accounts GDP or consumption per capita should get most of the weight, although possibly a little less weight when combined with income surveys than when combined with consumption surveys. Column 11 presents estimates for the entire sample of surveys, but using log national accounts consumption as the national accounts measure whenever the matching survey is a consumption survey, with the same results.

Currently, we treat each observation as a single "survey experiment" in which new values of the measurement error in national accounts GDP per capita, survey means and lights are realized, and so, we weigh all observations equally. However, if we wish to use our estimates to compute facts about the world distribution of income, it may be useful to overweigh countries with greater population (and which affect the world distribution of income more), and underweigh countries that happen to have a large number of surveys. Therefore, in column 12 of Table II we perform our baseline analyses while weighting the observations proportionally to the average country population in all of its survey years, and proportionally to the reciprocal of the number of surveys that the country has conducted in the period 1992-2010. We obtain much the same results as we do in the baseline, except in the country-year fixed effects specification, in which the estimated weight of log national accounts GDP per capita, $\hat{\omega}_G$, decreases to 0.67.[22] Nevertheless, our

[21] We obtain this variable from the World Development Indicators (Household Final Consumption Expenditure), and supplement it when it is missing by multiplying the consumption share from the Penn World Tables by the WDI estimate of GDP per capita.

[22] It is important to note that while all of our coefficients are robust to the construction of the aggregate lights measure, this coefficient varies with the lights measure. In particular, if instead of constructing the aggregate lights measure by averaging the digital number, we average the digital number taken to some power (such as 3/2 or 5/2) the weight on log GDP per capita increases to 0.75 or 0.79 respectively. Since when we calibrate the exponent empirically in Column 6, we find that it should be no less than 5/2, we believe that the 0.67 estimate is a lower bound on the relative weight of log GDP per capita in the weighted specification.

inferences remain almost the same as in the baseline.

It is important to note that for all the rows and columns of Table II, we always fail to reject the null hypothesis $\omega_G^* = 1$, or that one should only use national accounts GDP per capita in the optimal lights-based proxy, and we always reject the null hypothesis $\omega_G^* = 0$, or that one should only use survey means in the optimal lights-based proxy. For all but 8 of the 48 specifications in this table, we reject the null hypothesis that $\omega_G^* = 0.5$, or that GDP per capita and survey means should receive equal weight in the optimal lights-based proxy for true income. The eight specifications where we fail to reject entail wide standard errors rather than small weights on the national accounts $\hat{\omega}_G$, and two of these specifications involve samples much smaller than the baseline, which helps explain the wide confidence intervals. Over all of these specifications, the value of $\hat{\omega}_G$ does not fall below 0.67, which would correspond to placing two-thirds of the weight on log national accounts GDP per capita, and only one-third of the weight on log household survey means.

4.4 Relative Weights by Subsample

Our model in Section 3 may have very different parameter values for some types of countries than it does for the world on average. For example, different regions might have statistical agencies following different procedures, or surveys of varying quality. Similar arguments can be made for different time periods of our analysis. If these differences indeed are present, we may be concerned that our key finding of $\omega_G^* = 1$ is driven by averaging over a suitable mix of heterogeneous regions and years, and that this average may fail to be representative of many observations. Therefore, in this section, we will re-estimate our model on subsets of countries and years in our sample. While our sample sizes necessarily shrink, preventing us from reaching as precise inferences as in Sections 4.2 and 4.3, we nevertheless are able to make meaningful statements about the pattern of the coefficients. Reassuringly, we find that our model is remarkably homogeneous across different regions and years, which suggests that our estimates in Table II are representative of most countries' relationships between national accounts, household surveys and lights.

The columns of Table III present the same statistics and specifications as in the first column of Table II, but each estimated for a particular subsample. Column 1 replicates the baseline estimates, which are computed for the developing world as a whole, in order to facilitate comparison of the subsample estimates to the overall results. Columns 2 through 5 present estimates of optimal weights on log GDP per capita and log survey means for Africa, Asia, Latin America and the post-Communist countries (former Soviet republics and the Soviet Union's satellite states in Eastern Europe) respectively. For all of these specifications, we fail to reject the null hypothesis that $\omega_G^* = 1$, or that we should place full weight on log GDP per capita, and for all but 2 of these specifications we reject the hypothesis that $\omega_G^* = 0$, or that we should place full weight on log survey means. For Africa, Asia and the post-Communist countries, we further find that the optimal relative weight on log GDP per capita, ω_G^*, should be no less than 0.80, and typically should be around or even above unity in every specification. Moreover, for 8 of the 12 specifications we estimate in these three columns, we can reject the null hypothesis that $\omega_G^* = 0.5$, which means that log GDP per capita should get more than half the weight.[23] For Latin America, we see that surveys should play a greater role, with the weight on log GDP per capita ω_G^* varying from 0.78 (in the no fixed effects specification) to 0.59 (in the country and year fixed effects specification). It is likely that Latin American surveys are somewhat superior to the household surveys of other countries, although the difference is not enormous.

A particular worry may be that countries with high poverty counts may have different parameter values for the model in Section 3 than countries with low poverty counts. To address this concern, we divide countries into those with average poverty counts above the median (the more poor countries) and those with average poverty counts below the median (the less poor countries). Examples of the more poor countries are India, China, Nigeria, Bangladesh and the Democratic Republic of Congo, while examples of the less poor countries are Guinea-Bissau, Kazakhstan, Georgia, Botswana and Mongolia. Hence, countries can have more poor simply because they are large, as well as because they are poor. Columns 6 and 7 present our estimates for these subsamples of countries. To our reassurance, we see that the optimal weights for the more poor countries are similar to the optimal weights for the less poor countries, with estimated ω_G^* ranging from 0.78 to 1.18 across specifications. As usual, we can always reject the null hypothesis $\omega_G^* = 0$ and we always fail to reject the null hypothesis $\omega_G^* = 1$.[24]

Since different countries have statistical agencies of different quality, it may be the case that surveys dominate national accounts in places where the national accounts are poor. Robert Summers and Alan Heston, the authors of the Penn World Tables, have assigned subjective quality grades from A to D to the national accounts data of different countries.[25] On average, countries with higher grades have lower margins of error in their Penn World Tables GDP estimates, so this subjective classification appears to be reasonable. In our sample of developing countries, none earn a grade of A. We call all countries earning a grade of B or C the "good data countries" and we call all countries earning a grade of D "bad data countries."[26] The majority of countries (91 out of 123) are good data countries. In columns 8 and 9 we re-estimate our model for the good and the bad data countries separately. The point estimates on the optimal weight of log GDP

[23] In an additional 9th specification, we can reject the null that $\omega_G^* \leq 0.46$, which is effectively the hypothesis considered in Chen and Ravallion (2010).

[24] While we fail to reject the null hypothesis $\omega_G^* = 0.5$ for 2 of the 8 specifications in this section, we can reject the null $\omega_G^* = 0.46$, which is the effective null of the mixed method of Chen and Ravallion (2010) for all the eight specifications.

[25] We use the grades reported by Chen and Nordhaus (2011) to perform our classification. We assign a grade of D to all countries that are not graded by Summers and Heston, and which are given a grade of E by Chen and Nordhaus.

[26] Henderson, Storeygard and Weil (2012) use this terminology for a slightly different partition of countries.

per capita, ω_G^*, are all above 0.77, and all but three are above unity. As usual, we reject $\omega_G^* = 0$ and fail to reject $\omega_G^* = 1$ in all specifications. Our inferences for the optimal weights in good data countries are very similar to the baseline specifications. Perhaps more surprisingly, for the bad data countries we reject the null hypothesis $\omega_G^* = 0.5$ for three out of four specifications, suggesting that even in the countries with the worst national accounts statistics, the national accounts are still overwhelmingly superior to household surveys in measuring developing world living standards.

Finally, Columns 10 through 12 present estimates of optimal weights for the time periods 1992-1997, 1998-2003, and 2004-2010.[27] Once again, for all of these specifications, we fail to reject the null hypothesis that $\omega_G^* = 1$, or that we should place full weight on log GDP per capita. For the time period 1992-1997, the optimal weights on log GDP per capita are around 0.9, with the confidence intervals rejecting the null hypothesis $\omega_G^* = 0.5$ and failing to reject the null hypothesis $\omega_G^* = 1$. For the period 1998-2003, the magnitudes of the estimate of ω_G^* are typically around unity or higher (except in the country and year fixed effects specification where $\hat{\omega}_G = 0.59$), but their standard errors explode when country fixed effects are included, so for those two specifications we cannot reject any reasonable null hypothesis. Finally, for the period 2004-2010, our estimates of ω_G^* are typically lower and range from 0.63 to 1.15, indicating that surveys have improved over time. However, even for the 2004-2010 specifications, we can reject the null hypothesis that $\omega_G^* = 0$, or that we should place full weight on log survey means.

5 Application: Estimates of Poverty for the Developing World

One area in which the debate between national accounts and household surveys has direct implications is the estimation of world poverty. Researchers estimating poverty use data on within-country inequality from household surveys (because there is no alternative source of data) and combine it with an estimate for the mean of each country's income distribution. Some researchers (Chen and Ravallion 2001, 2004, 2010) use the mean of the surveys (that is, they just use household survey data). Other researchers (Bhalla 2002, Sala-i-Martin 2006, Pinkovskiy and Sala-i-Martin 2009, 2014) instead use GDP per capita or consumption per capita from the national accounts.[28] Since the level and the growth rate of survey means and GDP per capita are very different – survey means are lower and grow much slower, as we have seen in Section 4 – researchers using household surveys find much higher poverty and much slower poverty declines than do researchers using national accounts. Pinkovskiy and Sala-i-Martin (2009) and Dhonge and Minoiu (2010)

[27] We chose this division of the period 1992-2010 to have roughly equal numbers of years in each subperiod, and to isolate the years associated with the Asian financial crisis and its aftermath (1998-2003) in a single period.

[28] Researchers trying to measure global inequality also do not agree on whether to use national accounts or household surveys to measure the means of country income distributions. For example, Bourguignon and Morrisson (2002), Sala-i-Martin (2006), Pinkovskiy and Sala-i-Martin (2009, 2014) and Bourguignon (2015) use national accounts GDP per capita, while Milanovic (2005), Anand and Segal (2008), and Lackner and Milanovic (2013) use or advance arguments in favor of household surveys.

show that the choice of which series to use for the mean of each country's distribution is the driving factor for the differences between the two sets of estimates. Hence, for the poverty literature it is crucial to know whether one should use national accounts GDP per capita, survey means or a combination of the two.

Table IV shows poverty estimates for the sample of countries and years that is considered in this paper. For each country and year, we construct an estimate of the poverty rate by taking the integral below a specified poverty line of the lognormal distribution whose mean is the GDP per capita or the household survey mean of that country in that year. We consider poverty lines of \$1.25 a day (the official \$1-a-day poverty line used by the World Bank), and \$2.50 a day (twice the World Bank's poverty line). Row 1 presents \$1 a day poverty estimates using survey means, which show that poverty declined from 42% in 1992 to 20% in 2010. Row 2 presents \$1 a day poverty estimates computed using the same methodology, but using GDP per capita for the mean of the income distribution. These estimates show that poverty declined from 9.4% to 3.7%. Row 3 presents \$2 a day poverty estimates using survey means, which decline from 71% to 48%. Row 4 presents \$2 a day poverty estimates using GDP per capita, which decline from 31% to 11%. We observe that the survey-based estimates are very similar to the ones presented for the developing world by Chen and Ravallion (2010), while the GDP per capita-based estimates are similar to those of Pinkovskiy and Sala-i-Martin (2009).[29] It is important to note both the dramatic differences in levels between GDP and survey-based estimates, as well as the more rapid (in the case of the \$2 a day estimates, much more rapid) rate of reduction of poverty if measured using GDP per capita.[30]

Implicitly, both groups of researchers are using a proxy for the mean of the true income distribution of the form

$$z_{i,t} = \psi \left(\omega y^G_{i,t} + (1-\omega) y^S_{i,t} \right) \tag{24}$$

Researchers using household survey means implicitly assume that $\psi = 1$, and $\omega = 0$.[31] Researchers using national accounts GDP per capita implicitly assume $\psi = 1$, and $\omega = 1$. It has been acknowledged in the literature that intermediate cases can exist: for example, Karshenas (2003) and Chen and Ravallion (2010) consider $\psi = 1$, and $\omega = 0.5$. In this paper, we do not question the assumption that $\psi = 1$, but we provide some insight on what ω should be. Specifically, we find that ω should be much closer to 1 than to 0, or 0.5. Therefore, under the implicit assumptions that the literature has been making, we should place

[29] The differences with Pinkovskiy and Sala-i-Martin (2009) come from using a different dataset for GDP (the World Development Indicators vs. the Penn World Tables version 6.2), and from the estimates in this paper excluding the population of developed countries from the denominator.

[30] Very similar differences would obtain if national accounts consumption per capita were used in lieu of GDP per capita.

[31] Strictly speaking, the proxy should also include an intercept, which we have been treating as partialled-out in our analysis of Section 3. Hence, the proxy equation should read

$$z_{i,t} = \alpha + \psi \left(\omega y^G_{i,t} + (1-\omega) y^S_{i,t} \right)$$

Both researchers using survey means and researchers using national accounts assume that $\alpha = 0$.

more confidence in the poverty estimates in rows 2 and 4 (the ones based on GDP per capita) than in rows 1 and 3 (the ones based on survey means), which means that poverty is lower and is falling faster than the official estimates indicate. [32]

6 Effects of Including Lights in the Proxy

Up until this paper, the literature has used the lights variable as a way to improve GDP by including it directly as a right hand-side variable. Their version of our proxy equation (4) is

$$z_{i,t} = \gamma_L y_{i,t}^L + \gamma_G y_{i,t}^G \tag{25}$$

(Henderson, Storeygard and Weil 2012, Chen and Nordhaus 2011). We have instead used nighttime lights as an impartial refree to help us obtain the optimal weights on national accounts GDP per capita and survey means. A natural extension of our approach and of the nighttime lights literature is to see what happens when we augment our formula (4) for the optimal proxy to include lights directly, that is:

$$z_{i,t} = \gamma_L y_{i,t}^L + \gamma_G y_{i,t}^G + \gamma_S y_{i,t}^S \tag{26}$$

Now, we want to compute the vector of weights (γ_L,γ_G, γ_S), which minimizes the mean squared error

$$(\gamma_L^*, \gamma_G^*, \gamma_S^*) = \arg \min_{\gamma_G, \gamma_S} E\left(\left(y_{i,t}^* - \gamma_L y_{i,t}^L - \gamma_G y_{i,t}^G - \gamma_S y_{i,t}^S \right)^2 \right) \tag{27}$$

subject to the constraint that the proxy be unbiased, that is:

$$E\left(\gamma_L^* y_{i,t}^L + \gamma_G^* y_{i,t}^G + \gamma_S^* y_{i,t}^S | y_{i,t}^* \right) = y_{i,t}^* \tag{28}$$

Using the same steps as in Section 3, we can reformulate our optimization problem as

$$(\gamma_G^*, \gamma_S^*) = \arg \min_{\gamma_G, \gamma_S} \left\{ \gamma_L^2 var\left(\varepsilon_{i,t}^L\right) + \gamma_G^2 var\left(\varepsilon_{i,t}^G\right) + 2\gamma_G \gamma_S cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right) + \gamma_S^2 var\left(\varepsilon_{i,t}^S\right) \right\} \tag{29}$$

subject to

$$\gamma_L^* \beta^L + \gamma_G^* \beta^G + \gamma_S^* \beta^S = 1 \tag{30}$$

[32] A concern in interpreting the national accounts-based poverty rates could be that surveys underestimate not only the mean of the income distribution, but also inequality, and that this underestimation is highest for rich people in the developing world. In our working paper (Pinkovskiy and Sala-i-Martin 2014) we present evidence to alleviate this concern; namely, that accounting for survey misreporting in top income shares does not change our results substantially. We also show in Section 7 that GDP per capita correlates well with obvious indicators of the quality of life of the poor even after controlling for survey means. Given the scope of this paper, we leave this question for future research.

Solving this problem with traditional constrained optimization techniques, we obtain the following equation for $\frac{\gamma_G^*}{\gamma_S^*+\gamma_G^*}$:

$$\frac{\gamma_G^*}{\gamma_G^*+\gamma_S^*} = \frac{\beta^G var\left(\varepsilon_{i,t}^S\right) - \beta^S cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right)}{\beta^G var\left(\varepsilon_{i,t}^S\right) + \beta^S var\left(\varepsilon_{i,t}^G\right) - \left(\beta^G + \beta^S\right) cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right)} \tag{31}$$

Notice that this is exactly the same equation as Equation (11) in Section 3! Hence, adding nighttime lights to our proxy does not change the optimal weight on log GDP per capita relative to the optimal weight on log survey means. This is intuitive from equations (29) and (30), because the optimal weight on lights enters additively into both the objective and into the constraint of our optimization problem, and therefore, does not change the first order conditions for the optimal weights on GDP and surveys γ_G^* and γ_S^*. Additionally, including nighttime lights in the proxy has no bearing on our analysis of the regression equation (12), and therefore does not affect our conclusion that we can estimate the ratio of the weights $\omega_G^* = \frac{\gamma_G^*}{\gamma_G^*+\gamma_S^*}$ by the ratio of regression coefficients $\hat{\omega}_G = \frac{b^G}{b^G+b^S}$, so equation (13), the central result of our mathematical Section 3 and the core of our econometric approach in Section 4 remains valid. So all the tables that we have shown so far would still reflect the relative optimal weights of national accounts GDP per capita and survey means. Including lights has implications only for how to weigh lights relative to the optimal combination of national accounts GDP per capita and survey means, not on what this optimal combination of national accounts GDP per capita and survey means is.

However, it still might be interesting to look at what is the weight that nighttime lights should get in the optimal proxy relative to the optimal combination of national accounts GDP per capita and survey means ($\frac{\gamma_L^*}{\gamma_G^*+\gamma_S^*}$). The first-order conditions imply that:

$$\frac{\gamma_L^*}{\gamma_G^*+\gamma_S^*} = \frac{\beta^L}{var\left(\varepsilon_{i,t}^L\right)} \frac{\left(var\left(\varepsilon_{i,t}^G\right) var\left(\varepsilon_{i,t}^S\right) - cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right)^2\right)}{\left(\beta^G var\left(\varepsilon_{i,t}^S\right) + \beta^S var\left(\varepsilon_{i,t}^G\right) - \left(\beta^G + \beta^S\right) cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right)\right)} \tag{32}$$

Under Assumptions A1 and A2 we do not have enough moments in the data to estimate this equation. Therefore, we follow the approach of Henderson, Storeygard and Weil (2012) and make some additional assumptions. Specifically, we assume that

$$\beta^G = 1 \text{ and } var\left(\varepsilon_{i,t}^G\right) = \phi, \text{ which is known} \tag{L}$$

This assumption brings our model to be exactly the same as in Henderson, Storeygard and Weil (except that we also include surveys in our proxy). The content of Assumption L is that log GDP per capita is an unbiased indicator of log true income, which is not unreasonable to assume following our results in

Section 4, and that the margin of error of log GDP per capita is known, which is also not unreasonable to assume in light of the literature on the errors of GDP (Johnson et al. 2011, Chen and Nordhaus 2011). Using Assumption L we can readily identify $var\left(y^*_{i,t}\right)$

$$var\left(y^*_{i,t}\right) = var\left(y^G_{i,t}\right) - \phi \tag{33}$$

From here, we can identifiy the following variances, covariances and slope parameters

$$cov\left(y^L_{i,t}, y^G_{i,t}\right) = \beta^L var\left(y^*_{i,t}\right) \Rightarrow \beta^L = cov\left(y^L_{i,t}, y^G_{i,t}\right) / var\left(y^*_{i,t}\right) \tag{34}$$

$$var\left(y^L_{i,t}\right) = \left(\beta^L\right)^2 var\left(y^*_{i,t}\right) + var\left(\varepsilon^L_{i,t}\right) \Rightarrow var\left(\varepsilon^L_{i,t}\right) = var\left(y^L_{i,t}\right) - \left(\beta^L\right)^2 var\left(y^*_{i,t}\right) \tag{35}$$

$$cov\left(y^L_{i,t}, y^S_{i,t}\right) = \beta^L \beta^S var\left(y^*_{i,t}\right) \Rightarrow \beta^S = cov\left(y^L_{i,t}, y^S_{i,t}\right) / \left(\beta^L var\left(y^*_{i,t}\right)\right) \tag{36}$$

$$var\left(y^S_{i,t}\right) = \left(\beta^S\right)^2 var\left(y^*_{i,t}\right) + var\left(\varepsilon^S_{i,t}\right) \Rightarrow var\left(\varepsilon^S_{i,t}\right) = var\left(y^S_{i,t}\right) - \left(\beta^S\right)^2 var\left(y^*_{i,t}\right) \tag{37}$$

$$cov\left(y^G_{i,t}, y^S_{i,t}\right) = \beta^S var\left(y^*_{i,t}\right) + cov\left(\varepsilon^G_{i,t}, \varepsilon^S_{i,t}\right) \Rightarrow cov\left(\varepsilon^G_{i,t}, \varepsilon^S_{i,t}\right) = cov\left(y^G_{i,t}, y^S_{i,t}\right) - \beta^S var\left(y^*_{i,t}\right) \tag{38}$$

Note that all of the parameters appearing on the right hand-side of equation (32) are identified through equations (34)-(38) so the ratio $\frac{\gamma^*_L}{\gamma^*_G + \gamma^*_S}$ can be estimated, even though we have no left hand-side variable to run a regression like (12) because lights is now an explanatory variable.

We can estimate the ratio $\frac{\gamma^*_L}{\gamma^*_G + \gamma^*_S}$ in the no fixed effects specification similar to the first cell of Table II. We assume that the standard deviation of the error term of log GDP per capita is equal to 0.21 log points, which is the average of the margins of error in our sample based on the quality ranking of Chen and Nordhaus (2011). This means that the variance of the error of national accounts GDP per capita, ϕ, should be equal to 0.044. In this specification, we obtain that the optimal weight on nighttime lights is 0.111, with a 95% confidence interval of $(0.056, 0.166)$. Chen and Nordhaus (2011) report different estimates of optimal weights on nighttime lights for countries with different quality grades in the Penn World Tables. Given that the overwhelming majority of countries in our sample (86 out of 123) receive a quality grade of C (which stands for moderate data quality), we should compare our estimate to the cross-section weight for grade C countries reported in Chen and Nordhaus (2011). This weight is 0.0738, which is within the confidence interval for our own estimate.[33]

7 Why Do the Surveys and National Accounts Diverge?

We started our paper by pointing out that GDP per capita and survey means diverge. We have shown that nighttime lights are much better correlated with GDP per capita than with survey means. Our implicit reading of this fact has been that both nighttime lights and GDP reflect some notion of well-being better than surveys do. However, this need not be the case. It could be that nighttime lights and GDP are both correlated, but do not reflect things that lead to people's well-being, while surveys do. For example, national accounts and nighttime lights could be reflecting potentially wasteful military spending or other unproductive investment, and may have little to do with the living standards of most people, while surveys might reflect consumption of necessities. Just as all that glitters is not gold, so all that glistens need not be well-being.

In this section, we explore which of the two above hypotheses is correct. The obvious way to do this is to introduce additional widely accepted measures of quality of life and living standards, such as life expectancy, fertility, access to sanitation and safe water and primary education, and look at whether national accounts GDP or household survey means better correlate with them. Unfortunately, these variables are typically collected in much the same way as surveys are conducted, which means that their measurement errors will be correlated with the measurement errors of surveys. However, the errors in the quality of life variables are much more likely to be correlated with the errors in household surveys, rather than with the errors in the business surveys that underlie national accounts because these quality-of-life variables tend to come from household surveys, oftentimes the same surveys that collect data on household income and consumption. Thus, for life expectancy at birth, "complete vital registration systems are not common in developing countries. Therefore estimates of life expectancy must be derived from sample surveys or by applying indirect estimation techniques to registration, census, or survey data. Survey data are subject to recall error..." Similarly, the FAO writes that the depth of the food deficit is "computed from national household surveys where they are available, which is the case for a wide sub-sample of the monitored countries." The WHO determines access to improved sanitation facilities "based on national censuses and nationally representative household surveys," where "the coverage rates for water and sanitation are based

[33] Both our estimates and those of Chen and Nordhaus (2011) yield a lower optimal weight on nighttime lights than Henderson, Storeygard and Weil (2012), who get an optimal weight of 0.5 for the countries with the poorest data quality. However, this is because Henderson, Storeygard and Weil (2012) look at countries of much worse data quality (countries receiving the lowest data quality ratings from the World Bank, which are comparable to countries rated D or below in the Penn World Tables and Chen and Nordhaus 2011). Moreover, Henderson, Storeygard and Weil (2012) look at estimating long-run growth rates only, while both we and Chen and Nordhaus (2011) estimate the cross-sectional or the panel distribution of output. Therefore, both this paper and Chen and Nordhaus (2011) estimate different objects than Henderson, Storeygard and Weil (2012).

on information from service users on the facilities their households actually use rather than on information from service providers." Therefore, we should expect our estimates of the correlations between quality-of-life measures, GDP per capita and household surveys to be biased in favor of finding too large a correlation with surveys and too small a correlation with GDP per capita.

We consider nine different indicators of well-being, all from the World Development Indicators. These are 1) log life expectancy in years, 2) the negative of the log of the fertility rate, 3) the negative of the log of the fertility rate among women aged 15 to 19, 4) the negative log of the food deficit in kilograms among people failing basic nutritional needs, 5) the negative of the log of the fraction of pregnant women suffering from anemia, 6) the log of the fraction of people with access to improved sanitation, 7) the log of the fraction of people with access to a safe water source, 8) the log of the fraction of primary school-aged children attending school and 9) the log of the female literacy rate. It is clear that all of these indicators unambiguously reflect increased welfare in developing countries; in the language of Young (2012), they are "patently obvious" indicators of good outcomes that policymakers care about and want to encourage.

Table V provides regressions of each of these indicators on log national accounts GDP per capita and log household survey mean income on their own (panel 1) and with (panel 2) country fixed effects, so as to analyze both level and growth rate variation. Row 1 starts off by reproducing parts of Column 4 from Table I – the bivariate regression of lights on log national accounts GDP per capita and log survey means – and subsequent rows change the dependent variable. In results not reported, all of these development indicators have statistically significant univariate relations with both GDP per capita and survey means individually, with and without country fixed effects. However, when both GDP per capita and survey means are included in the same regression, the coefficient on log GDP per capita is significant at least at 10% for all measures of well-being (and at 5% for all measures but one, primary schooling without country fixed effects). On the other hand, the coefficient on survey means is always smaller than the coefficient on log GDP per capita and fails to be significant in 4 of the specifications without country fixed effects and all but one specification with country fixed effects.[34] For example, when we regress the negative of the log of fertility per 100,000 people on log GDP per capita and log household survey mean (in Panel 1 and Column 3), the coefficient on log GDP per capita is a very significant 0.371 (s.e. = 0.087) and the coefficient on log household survey mean is an insignificant 0.034 (s.e. = 0.101). Therefore, we can conclude that in our case, what glitters is indeed gold, and national accounts (as well as nighttime lights) measure something that is fundamentally connected to people's well-being.

It is natural to ask why exactly surveys perform worse at measuring true income than do national

[34] Owing to the relatively low t-statistics on the partial coefficient on national accounts (compared to the t-statistic when nighttime lights are the dependent variable), the ratios of the two effects have nonstandard distributions, and therefore, we do not present them.

accounts. Our preceding analysis gives us a clue. It must be the case that survey questions on income and consumption (or features of the implementation of specifically the household surveys that ask about income and consumption) must have problems of their own, distinct from other survey questions and distinct from questions about elementary quality-of-life indicators. Otherwise, survey mean incomes and consumption levels would have reflected the income and consumption of the people answering questions about health and education, which should have lead them to have strong explanatory power for these measures of welfare.

The next two panels of Table V provide further information as to what may be going on in the surveys. Panels 3 and 4 present regressions of the difference in the logs of GDP per capita and household survey means (a measure of the bias of household survey means) on the quality-of-life measures used in Panels 1 and 2, as well as on nighttime light intensity. We see that the difference between log GDP per capita and log household survey means increases with every one of these measures in levels, and for many of these measures (in particular, the ones connected to health and literacy, though not food, sanitation or primary school attendance) in growth rates. In particular, it is useful to see in Column 1 and Panel 4 that the difference increases in the growth rate of nighttime lights (hence, of true income, since the errors in nighttime lights are independent of errors in national accounts GDP per capita or survey means). Therefore, countries with higher and growing well-being tend to suffer from progressively greater mismeasurement of income by surveys.

A possible explanation for this phenomenon may be that survey questions on income and consumption are notoriously complicated and vary in important ways across surveys even within the same country, while survey questions on life expectancy, fertility, access to sanitation and education are straightforward. Deaton (2005) describes the many detailed questions that a respondent needs to answer in order to generate an estimate of his or her consumption, as well as the extent to which the resulting estimate can be affected by technical features of the survey like the recall period. It takes a lot of time and effort for respondents to provide accurate answers to income and consumption survey questions, much more so than to questions about health, fertility, and other obvious measures of well-being. Since people generally have higher opportunity costs of time in richer and faster-growing countries, they are therefore likely to answer income and consumption questions relatively inaccurately compared to their answers to questions about obvious measures of well-being in richer and faster-growing countries than in poorer and slower-growing ones. In addition to this opportunity cost, people's income sources and consumption baskets in richer and faster-growing countries are more varied, and therefore harder to recall at once and report accurately to a survey administrator. Therefore, survey questions on income and consumption fail to capture some income and growth, leading to a discrepancy with national accounts GDP per capita, and with true income, that widens over time.

8 Conclusion

A major controversy in the literature on measuring developing world living standards is whether to use national accounts or household surveys. The discrepancies between these two sources are strikingly relevant because household survey means tend to be smaller and grow more slowly than national accounts GDP per capita. The implications are particularly important in measuring developing world poverty rates.

In this paper, we show that we can use a third variable as an impartial referee to tell how much weight to give to GDP per capita versus survey means in estimating developing world living standards. The requirement on this variable is that its error in measuring true income should be orthogonal to the measurement errors of both national accounts and household surveys. We argue that satellite-recorded nighttime lights satisfy this requirement. Although nighttime lights are a very noisy proxy for living standards, their measurement error comes from climatic disturbances to having a clear view of nighttime lights as well as from satellite aging. There is no reason to believe that this error is related to the primary sources of errors in national accounts and household surveys, which are survey nonresponse and faulty assumptions by statistical agencies.

Using the nighttime lights data, we can compute the optimal linear combination of GDP per capita and household survey means to proxy for true income. We find that in this combination, we want to give nearly all the weight to GDP per capita. This conclusion is robust to multiple alternative specifications and conceptual checks. It also holds not just in the aggregate, but in virtually every subsample of the data: different regions of the world, different time periods, and different groups of countries by poverty and by the quality of the national accounts systems.

We believe that the reason for the discrepancy between survey means on the one hand and GDP per capita and nighttime lights on the other hand is that surveys fail to capture larger and larger fractions of income in richer and faster-growing countries. Some reasons for why this could be taking place are that people have higher opportunity costs of time when incomes or growth rates are higher, or that people's income and consumption streams are more complicated to report in richer and faster-growing environments. We show that people tend to answer straightforward survey questions about their livelihood, such as whether their children are healthy and going to school, in a way that correlates much better with GDP per capita than with survey means of income and consumption. Therefore, it is likely that household survey income and consumption modules are faulty, perhaps by being overly complicated.

One of the areas in which the debate between national accounts and household surveys is crucial is the poverty literature. In April 2013, the World Bank set a new goal of reducing the fraction of people living on less than $1.25 a day to less than 3% by 2030. It stated that "Reducing the global extreme poverty rate to no more than 3 percent in 2030 is... difficult but achievable" (World Bank, 2013b). According to the household surveys, poverty for the world as a whole in 2010 is estimated to be 20%, making the World Bank's goal indeed challenging. However, according to the national accounts, the poverty rate in 2010 is only 3.7%, which is extremely close to the 3% target. So long as national accounts and nighttime lights adequately measure the well-being of the poor (which Section 7 suggests they do), it might be the case that by the time the World Bank goal was set in 2013, it had already been achieved.

9 Author Affiliations

Maxim Pinkovskiy, Federal Reserve Bank of New York

Xavier Sala-i-Martin, J. and M. Grossman Professor of Economics, Columbia University and NBER

References

[1] **Adcock, R. J.** "A Problem in Least Squares." *Analyst.* 5 (1878), pp. 53-4.

[2] **Alesina, Alberto F., Stelios Michalopoulos, and Elias Papaioannou**. "Ethnic Inequality." NBER WP #18512, 2012.

[3] **Ashraf, Quamrul, and Oded Galor**. "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review* 103 (2013), no. 1: 1-46.

[4] **Barro, Robert J.** "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics* 106 (1991), no. 2: 407-443.

[5] **Barro, Robert J.** "Human Capital and Growth in Cross-Country Regressions." *Swedish Economic Policy Review* 6 (1999), no. 2: 237-277.

[6] **Barro, Robert J., and Xavier Sala-i-Martin**. "Convergence." *Journal of Political Economy* 100 (1992), no. 2: 223-251.

[7] **Bhalla, Surjit S.** *Imagine there's no country: Poverty, inequality, and growth in the era of globalization.* Washington, D.C.:Penguin. 2002.

[8] **Bourguignon, Francois, and Christian Morrisson**. "Inequality among World Citizens: 1820-1992." *American Economic Review* 92 (2002), no. 4: 727-744.

[9] **Bourguignon, Francois**. *The Globalization of Inequality.* Princeton University Press. 2015

[10] **Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT)**. Gridded Population of the World, Version 3 (GPWv3). Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw. 2005.

[11] **Chen, Shaohua, and Martin Ravallion**. "How Did the World's Poorest Fare in the 1990s?" *Review of Income and Wealth* 47 (2001), no. 3: 283-300.

[12] **Chen, Shaohua, and Martin Ravallion**. "The Developing World Is Poorer Than We Thought, but No Less Successful in the Fight against Poverty." *Quarterly Journal of Economics* 125 (2010), no. 4: 1577-1625.

[13] **Chen, Shaohua, Martin Ravallion and Prem Sangraula**. *Povcalnet*, http://iresearch.worldbank.org/PovcalNet/index.htm. Accessed October 1, 2013.

[14] **Chen, Shaohua, and Martin Ravallion**. "How Have the World's Poorest Fared since the Early 1980s?." *World Bank Research Observer* 19 (2004), no. 2: 141-169.

[15] **Chen, Xi, and William D. Nordhaus**. "Using Luminosity Data as a Proxy for Economic Statistics" Proceedings of the National Academy of Sciences. 2011

[16] **Ciccone, Antonio and Marek Jarocinski**. "Determinants of Economic Growth: Will Data Tell?" *American Economic Journal: Macroeconomics* 2 (2010):4, 222-246.

[17] **Deaton, Angus**. "Counting the World's Poor: Problems and Possible Solutions." *World Bank Research Observer* 16 (2001), no. 2: 125-147.

[18] **Deaton, Angus**. "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)." *Review of Economics and Statistics* 87 (2005), no. 1: 1-19.

[19] **Deaton, Angus**. "Price Indexes, Inequality, and the Measurement of World Poverty." *American Economic Review* 100 (2010), no. 1: 5-34.

[20] **Dhongde, Shatakshee and Camelia Minoiu**. "Global Poverty Estimates: Present and Future." ECINE Working Paper Series, ECINEQ-WP 2010 #281

[21] **Doll, Christopher N.H.** "CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications." Manuscript. 2008

[22] **Doll, Christopher N.H.**, **Jan Peter Muller and Jeremy G. Morley**. "Mapping Regional Economic Activity from Night-Time Light Satellite Imagery." *Ecological Economics* 57 (2006): 75-92

[23] **Elvidge, Christopher D. and Kimberly E. Baugh, Eric A. Kihn, Herbert W. Kroehl, Ethan R. Davis**. "Mapping City Lights With Nighttime Data from the DMSP Operational Linescan System." *Photogrammetric Engineering & Remote Sensing* 63 (1997), no. 6: 727-734

[24] **Elvidge, Christopher D., Kimberly E. Baugh, John B. Dietz, Theodore Bland, Paul C. Sutton and Herbert W. Kroehl.** "Radiance Calibration of DMSP-OLS Low-Light Imaging Data of Human Settlements." *Remote Sensing of Environment*, 68 (1999) no. 1:77-88.

[25] **Elvidge, Christopher D., Kimberly E. Baugh, Sharolyn. J. Anderson, Paul. C. Sutton, and Tilottama Ghosh**. "The Night Light Development Index (NLDI): A Spatially Explicit Measure of Human Development from Satellite Data." *Social Geography* 7 (2012), 23–35.

[26] **Fuller, Wayne**. *Measurement Error Models.* New York: John Wiley & Sons. 1987.

[27] **Ghosh, Tilottama, Rebecca L. Powell, Christopher D. Elvidge, Kimberly E. Baugh, Paul C. Sutton and Sharolyn Anderson**. "Shedding Light on the Global Distribution of Economic Activity." *The Open Geography Journal* 3 (2010), 148-161.

[28] **Griliches, Zvi**. "Economic Data Issues," in *Handbook of Econometrics, Volume III.* Z. Griliches and M. D. Intriligator, eds. Amsterdam, New York and Oxford: North-Holland, pp. 1465–514. 1986.

[29] **Hausman, Jerry A.** "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left." *Journal of Economic Perspectives*, Vol. 15 (2001), No. 4, pp. 57-67.

[30] **Henderson, J. Vernon, Adam Storeygard, and David N. Weil**. "Measuring Economic Growth from Outer Space." NBER Working Paper #15199, 2009.

[31] **Henderson, J. Vernon, Adam Storeygard, and David N. Weil**. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2012), no. 2: 994-1028.

[32] **Johnson, Simon, William Larson, Chris Papageorgiou, and Arvind Subramanian**. "Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates." *Journal of Monetary Economics* 60 (2013), no. 2: 255-274.

[33] **Karshenas, Massoud**, "Global Poverty: National Accounts Based versus Survey Based Estimates," *Development and Change* 34 (2003) no. 4: 683-712.

[34] **Baugh, Kimberly E., Christopher D. Elvidge, Tilottama Ghosh and Daniel Ziskin**. "Development of a 2009 Stable Lights Product using DMSP-OLS Data." *Proceedings of the 30th Asia-Pacific Advanced Network Meeting.* 2009.

[35] **Korinek, Anton, Johan Mistiaen and Martin Ravallion**. "Survey Nonresponse and the Distribution of Income." *Journal of Economic Inequality* 4 (2006) no. 2: 33–55.

[36] **Lackner, Christoph and Branko Milanovic**. "Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession." World Bank Policy Research Working Paper #6719, 2013.

[37] **Luxembourg Income Study (LIS) Database, Key Figures**, http://www.lisdatacenter.org (multiple countries; [1967-2010]). Luxembourg: LIS.

[38] **Mankiw, N. Gregory, David Romer, and David N. Weil**. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107 (1992), no. 2: 407-437.

[39] **Michalopoulos, Stelios, and Elias Papaioannou**. "Pre-colonial Ethnic Institutions and Contemporary African Development." *Econometrica* 81 (2013), no. 1: 113-152.

[40] **Michalopoulos, Stelios, and Elias Papaioannou**. 2014. "National Institutions and Sub-national Development in Africa." *Quarterly Journal of Economics* 129 (2014) no. 1:151-213.

[41] **Milanovic, Branko**. *Worlds Apart: Measuring International and Global Inequality.* Princeton and Oxford: Princeton University Press. 2005.

[42] **Pinkovskiy, Maxim L**. "World Welfare Is Rising: Estimation Using Nonparametric Bounds on Welfare Measures." *Journal of Public Economics* 97 (2013), pp. 176-95

[43] **Pinkovskiy, Maxim L.** "Economic Discontinuities at Borders: Estimation Using Satellite Data of Lights at Night." mimeo, Federal Reserve Bank of New York. 2014.

[44] **Pinkovskiy, Maxim L. and Xavier Sala-i-Martin**. "Parametric Estimations of the World Distribution of Income." NBER Working Paper #15433, 2009.

[45] **Pinkovskiy, Maxim L. and Xavier Sala-i-Martin**. "Africa is on Time." *Journal of Economic Growth* 19 (2014) :311-338.

[46] **Sala-i-Martin, Xavier**. "The Disturbing 'Rise' of Global Income Inequality." NBER Working Paper #8904, 2002a.

[47] **Sala-i-Martin, Xavier**. "The World Distribution of Income (estimated from Individual Country Distributions)." NBER Working Paper #8933, 2002b.

[48] **Sala-i-Martin, Xavier**. "The World Distribution of Income: Falling Poverty and . . . Convergence, Period." *Quarterly Journal of Economics* 121 (2006), no. 2: 351-397.

[49] **Sutton, Paul C., Christopher D. Elvidge and Tilottama Ghosh**. "Estimation of Gross Domestic Product at Sub-National Scales using Nighttime Satellite Imagery." *International Journal of Ecological Economics & Statistics* 8 (2007) :5-21

[50] **World Bank**. *World Development Indicators*. Washington, DC. 2013a.

[51] **World Bank**. "Ending Extreme Poverty and Promoting Shared Prosperity." http://www.worldbank.org/en/news/feature/2013/04/17/ending_extreme_poverty_and_promoting_shared_prosperi 2013b.

[52] **Young, Alwyn**. "The African Growth Miracle." *Journal of Political Economy* 120 (2012), no. 4, pp. 696-739.

10 Tables

Table I (I)

Baseline Regressions	(1)	(2)	(3)	(4)	(5)
	Log Lights OLS	Log Lights OLS	Log Surveys IV, Lights	Log Lights OLS	Log Lights OLS
No Fixed Effects					
Log GDP per Capita	1.160***		.779***	1.020***	.514***
	(.063)		(.027)	(.129)	(.124)
Log Survey Mean Income		1.286***		.184	.035
		(.081)		(.135)	(.113)
Log Electricity per Capita					.449***
					(.063)
R2	.72	.63	.81	.73	.81
Year Fixed Effects					
Log GDP per Capita	1.171***		.780***	1.003***	.521***
	(.065)		(.028)	(.130)	(.131)
Log Survey Mean Income		1.303***		.221	.060
		(.082)		(.135)	(.119)
Log Electricity per Capita					.442***
					(.064)
R2	.75	.65	.82	.75	.83
Country Fixed Effects					
Log GDP per Capita	.559***		.626***	.593***	.383**
	(.094)		(.140)	(.132)	(.178)
Log Survey Mean Income		.306***		-.049	-.032
		(.069)		(.095)	(.099)
Log Electricity per Capita					.252**
					(.124)
R2	.97	.96	.95	.97	.95
Country and Year Fixed Effects					
Log GDP per Capita	.661***		.330	.676***	.497**
	(.148)		(.312)	(.173)	(.200)
Log Survey Mean Income		.111		-.033	-.017
		(.093)		(.102)	(.112)
Log Electricity per Capita					.278**
					(.132)
R2	.98	.97	.96	.98	.97
Number of Obs.	701	701	701	701	617
Number of Clusters	123	123	123	123	92

Table I presents estimates for the regressions of log nighttime lights per capita on log national accounts GDP per capita and / or log survey mean income or consumption per capita, as described in Section 4.1. Column 3 presents results for a regression of log survey means on log GDP per capita instrumented with log nighttime lights per capita. Standard errors in parentheses are clustered by country. Data on nighttime lights from the NOAA, data on national accounts GDP from the World Development Indicators, and data on survey means is from Chen and Ravallion (2010).

Table II (II)

Weights in the Optimal Proxy: Robustness Checks

Dependent Variable is Log Light Intensity per Capita unless otherwise noted

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Baseline	Additional Covariates			Different Dep. Var.			Different NAS Variables				
		Elect Ricity	All Controls	Nonlinear Controls	Log Light Density	Calibrated Lights	Fraction Pop. Lit.	NA Cons umption	Income Surveys	Cons. Surveys	Match Concepts	Weighted
No Fixed Effects												
Log GDP per Capita	.84***	.93***	1.15***	.99***	1.11***	.81***	1.06***	.85***	1.01***	.89***	.88***	1.14***
	(.64,1.05)	(.52,1.3)	(.67,1.8)	(.59,1.5)	(.69,1.7)	(.54,1.1)	(.61,1.7)	(.65,1.1)	(.57,1.6)	(.57,1.2)	(.65,1.09)	(.70,1.5)
Log Survey Mean Income	.15	.06	-.15	.00	-.11	.18	-.06	.14	-.01	.10	.11	-.14
	(-.05,.35)	(-.38,.47)	(-.89,.32)	(-.58,.40)	(-.72,.30)	(-.15,.45)	(-.72,.38)	(-.10,.34)	(-.69,.42)	(-.26,.42)	(-.09,.34)	(-.56,.29)
Year Fixed Effects												
Log GDP per Capita	.81***	.89***	1.05***	.92***	1.07***	.80***	1.04***	.82***	.93***	.84***	.84***	1.13***
	(.63,1.03)	(.46,1.3)	(.57,1.7)	(.52,1.3)	(.65,1.6)	(.52,1.1)	(.59,1.6)	(.63,1.05)	(.45,1.5)	(.51,1.2)	(.64,1.07)	(.73,1.5)
Log Survey Mean Income	.18*	.10	-.05	.07	-.07	.19	-.04	.17	.06	.15	.15	-.13
	(-.03,.36)	(-.30,.53)	(-.73,.42)	(-.35,.47)	(-.67,.34)	(-.15,.47)	(-.68,.40)	(-.05,.36)	(-.52,.54)	(-.23,.48)	(-.07,.35)	(-.50,.26)
Country Fixed Effects												
Log GDP per Capita	1.09***	1.09**	1.18***	1.15***	.99***	1.08***	1.12***	1.02***	.92***	1.23***	1.04***	.96***
	(.77,1.3)	(.27,1.7)	(.79,1.4)	(.77,1.3)	(.73,1.2)	(.84,1.3)	(.77,1.7)	(.60,1.2)	(.60,1.2)	(.80,1.5)	(.72,1.3)	(.54,1.1)
Log Survey Mean Income	-.09	-.09	-.18	-.15	.00	-.08	-.12	-.02	.07	-.23	-.04	.03
	(-.36,.22)	(-.77,.72)	(-.47,.20)	(-.38,.22)	(-.21,.26)	(-.32,.15)	(-.79,.22)	(-.26,.39)	(-.27,.39)	(-.57,.19)	(-.31,.27)	(-.17,.45)
Country and Year Fixed Effects												
Log GDP per Capita	1.05***	1.03**	1.08***	1.05**	1.07***	1.12***	.98**	1.03***	.76**	1.14***	1.04***	.67***
	(.72,1.3)	(.35,1.4)	(.69,1.3)	(.64,1.2)	(.53,1.5)	(.83,1.4)	(.30,1.4)	(.40,1.4)	(.00,1.1)	(.79,1.7)	(.73,1.3)	(.47,1.08)
Log Survey Mean Income	-.05	-.03	-.08	-.05	-.07	-.12	.01	-.03	.23	-.14	-.04	.32
	(-.30,.27)	(-.45,.64)	(-.32,.30)	(-.26,.35)	(-.53,.46)	(-.42,.16)	(-.44,.69)	(-.49,.59)	(-.12,.99)	(-.71,.20)	(-.31,.26)	(-.08,.52)
Number of Obs.	701	617	565	565	701	701	160	701	263	438	701	701
Number of Clusters	123	92	87	87	123	123	82	123	43	99	123	123

Each column of Table II presents estimates of the relative weights of log GDP per capita and log survey means in the optimal lights-based proxy $z_{i,t}$ of the mean of the true income distribution. Block-bootstrapped 95% confidence intervals in parentheses. The baseline specification does not include covariate controls, and uses log aggregate lights per capita to measure light intensity. Column 2 controls for log electricity production per capita. Column 3 controls for log electricity production per capita, log total population, log % rural population, log % urban population, log area, latitude and longitude, the income share of the richest 10% and the income share of the poorest 50%, log consumption share, log capital formation as % of GDP, log shares of GDP in agriculture, manufacturing and services, log export share, log import share, log government expenditure share of GDP, log GDP per energy unit consumed and log oil rents. Column 4 includes the controls in column 3 as well as their squares. Column 5 replaces the dependent variable with log light density. Columns 6 and 7 replace the dependent variable with log calibrated lights per capita, where the calibration is done to optimize fit to LIS data on Mexican state incomes (LIS 2013) and log fraction of the country's population that resides in lit areas, respectively. Column 8 replaces log GDP per capita with log national accounts consumption per capita. Column 9 considers only the sample of income surveys. Column 10 considers only the sample of consumption surveys, and replaces log GDP per capita with log national accounts consumption per capita. Column 11 replaces log GDP per capita with log national accounts consumption per capita whenever the corresponding survey is a consumption survey. Column 12 weights all observations by average country population divided by the number of surveys for that country.

Table III

(III)

Weights in the Optimal Proxy: Subsamples

Dependent Variable is Log Light Intensity per Capita

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Baseline	Regions				Poverty		Grades		Years		
		Africa	Asia	Latin America	Post Comm.	Less Poor	More Poor	Good Data	Bad Data	1992 -1997	1998 -2003	2004 -2010
No Fixed Effects												
Log GDP per Capita	.84***	1.20***	1.32**	.84***	.80***	.79***	1.16***	.80***	1.03***	.88***	1.03***	.68***
	(.64,1.05)	(.74,1.9)	(.31,5.7)	(.31,1.6)	(.50,1.3)	(.51,1.1)	(.82,1.6)	(.58,1.03)	(.60,1.5)	(.59,1.1)	(.76,1.4)	(.46,1.01)
Log Survey Mean	.15	-.20	-.32	.15	.19	.20	-.16	.19*	-.03	.11	-.03	.31*
	(-.05,.35)	(-.93,.25)	(-4.7,.68)	(-.67,.68)	(-.36,.49)	(-.13,.48)	(-.66,.17)	(-.03,.41)	(-.55,.39)	(-.10,.40)	(-.46,.23)	(-.01,.53)
Year Fixed Effects												
Log GDP per Capita	.81***	1.12***	1.18**	.79***	.82***	.78***	1.09***	.77***	1.00***	.87***	1.05***	.63***
	(.63,1.03)	(.73,1.6)	(.18,4.1)	(.29,1.4)	(.54,1.3)	(.50,1.1)	(.73,1.5)	(.56,1.00)	(.69,1.3)	(.60,1.09)	(.78,1.4)	(.38,.96)
Log Survey Mean	.18*	-.12	-.18	.20	.17	.21	-.09	.22*	-.00	.12	-.05	.36***
	(-.03,.36)	(-.67,.26)	(-3.1,.81)	(-.49,.70)	(-.37,.45)	(-.10,.49)	(-.51,.26)	(-.00,.43)	(-.39,.30)	(-.09,.39)	(-.45,.21)	(.03,.61)
Country Fixed Effects												
Log GDP per Capita	1.09***	1.26*	1.13***	.74***	1.37**	1.18***	.96***	1.04***	1.47**	.89***	4.80	1.15***
	(.77,1.3)	(-1.1,5.0)	(.78,1.5)	(.47,1.1)	(.56,2.1)	(.68,1.5)	(.48,1.5)	(.69,1.3)	(.25,3.9)	(.62,1.04)	(-7.1,6.8)	(.99,1.3)
Log Survey Mean	-.09	-.26	-.13	.25	-.37	-.18	.03	-.04	-.47	.10	-3.80	-.15*
	(-.36,.22)	(-4.0,2.1)	(-.53,.21)	(-.16,.52)	(-1.1,.43)	(-.53,.31)	(-.59,.51)	(-.35,.30)	(-2.9,.74)	(-.04,.37)	(-5.8,8.1)	(-.38,.00)
Country and Year Fixed Effects												
Log GDP per Capita	1.05***	1.05*	1.01***	.59**	1.24***	1.14***	.90***	.99***	1.29***	.88***	.59	.82*
	(.72,1.3)	(-.01,3.9)	(.56,1.5)	(.07,1.07)	(.82,2.9)	(.56,1.4)	(.47,1.4)	(.63,1.3)	(.69,2.5)	(.67,1.01)	(-4.2,5.2)	(-.02,2.8)
Log Survey Mean	-.05	-.05	-.01	.40*	-.24	-.14	.09	.00	-.29	.11*	.40	.17
	(-.30,.27)	(-2.9,1.01)	(-.58,.43)	(-.07,.92)	(-1.9,.17)	(-.46,.43)	(-.41,.52)	(-.34,.36)	(-1.5,.30)	(-.01,.32)	(-4.2,5.2)	(-1.8,1.02)
Number of Obs.	701	114	119	234	234	406	295	614	87	165	234	302
Number of Clusters	123	41	29	25	28	61	62	91	32	88	98	103

Each column of Table III presents estimates of the relative weights of log GDP per capita and log survey means in the optimal lights-based proxy $z_{i,t}$ of the mean of the true income distribution. Block-bootstrapped 95% confidence intervals in parentheses. Each row corresponds to estimating the weights for a different subsample of the baseline sample: either restricting to observations in a specific region, to observations in a specific year range, to observations in countries with above or below median poverty count, or to observations with a specific set of Penn World Tables quality grades.

Table IV (IV)

Developing World Poverty Estimates:								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1992	2005	2006	2007	2008	2009	2010	Ratio 2010-1992
Panel I: $1/Day Poverty Rates								
(1) Survey Weight = 1 (CR 2010)	.421	.258	.247	.237	.227	.214	.205	.487
(2) GDP Weight = 1 (PSiM 2009)	.094	.050	.047	.043	.041	.039	.037	.400
Panel II: $2/Day Poverty Rates								
(3) Survey Weight = 1 (CR 2010)	.710	.553	.538	.524	.510	.490	.476	.670
(4) GDP Weight = 1 (PSiM 2009)	.313	.156	.143	.130	.123	.116	.109	.348

Poverty estimates are constructed by assuming that the income distribution in each country is lognormal, with the mean equal to the country's survey mean (Rows 1 and 3) or GDP per capita (Rows 2 and 4), and with the standard deviation of log income obtained by inverting the Gini coefficient. Survey means and within-country standard deviations of log income are interpolated or extrapolated whenever a survey is unavailable. The world poverty rate is then the population-weighted average fraction of each distribution below $1.25 per day for the first panel and below $2.50 per day for the second panel.

Table V

(V)

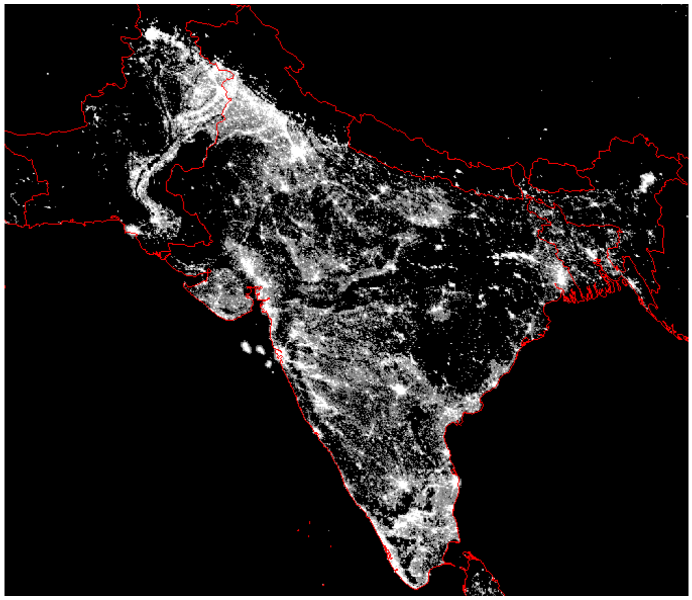
Quality-of-Life Indicators, National Accounts and Survey Means

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Log Lights per capita	Log Life Expectancy Years	Neg. Log Fertility per 100,000	Neg. Log Adolesc. Fertil. per 100,000	Neg. Log Food Deficit per capita	Log Frac Pregnant Anemic	Log Frac Access Sanitation	Log Frac Access Safe Water	Log Frac Primary School	Log Female Literacy Rate
QOL Measure on National Accounts and Survey Means: No Fixed Effects										
Log GDP per Capita	1.020***	.064***	.371***	.521***	1.033***	.125***	.405***	.137***	.059*	.215**
	(.129)	(.017)	(.087)	(.140)	(.259)	(.037)	(.084)	(.035)	(.032)	(.098)
Log Survey Mean Income	.184	.060***	.034	-.154	-.132	.086*	.141*	.075*	.028	.207*
	(.135)	(.021)	(.101)	(.176)	(.287)	(.044)	(.081)	(.039)	(.030)	(.119)
R2	.73	.58	.50	.24	.45	.46	.56	.55	.14	.59
QOL Measure on National Accounts and Survey Means: Country Fixed Effects										
Log GDP per Capita	.593***	.081***	.215***	.516***	.522**	.244***	.229***	.137***	.084**	.211***
	(.120)	(.011)	(.051)	(.085)	(.203)	(.040)	(.069)	(.036)	(.038)	(.063)
Log Survey Mean Income	-.049	.004	-.014	-.011	.366**	.039	.050	.033	.022	-.015
	(.086)	(.009)	(.043)	(.056)	(.161)	(.028)	(.036)	(.023)	(.036)	(.043)
R2	.21	.28	.14	.38	.18	.44	.20	.21	.04	.26
National Accounts-Survey Means Differential on QOL Measure: No Fixed Effects										
Log QOL Measure	.138***	.871***	.291***	.174***	.125***	.428***	.222***	.522***	.352***	.358***
	(.018)	(.212)	(.059)	(.036)	(.022)	(.099)	(.037)	(.110)	(.102)	(.071)
R2	.19	.08	.14	.11	.16	.08	.12	.10	.03	.19
National Accounts-Survey Means Differential on QOL Measure: Country Fixed Effects										
Log QOL Measure	.141**	.942**	.262**	.252***	-.007	.346**	.137	.225	.068	.482**
	(.057)	(.375)	(.107)	(.085)	(.031)	(.172)	(.084)	(.162)	(.134)	(.225)
R2	.03	.02	.02	.05	.00	.02	.00	.00	.00	.03
Number of Obs.	701	700	701	701	666	701	688	682	621	120
Number of Clusters	123	123	123	123	119	123	123	122	114	63

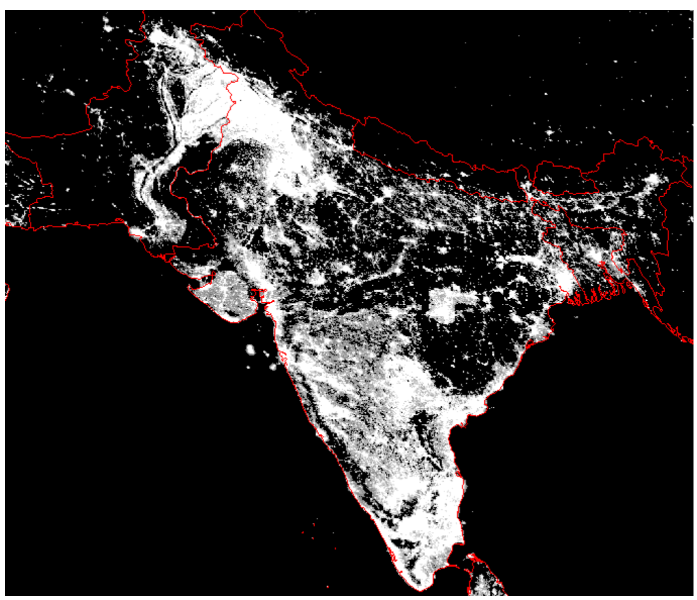
Each column in each panel of Table V presents coefficients from a regression of a proxy variable from the World Development Indicators onto log GDP per capita from the WDI, log household survey mean and (in the bottom panel) country fixed effects. All dependent variables are obtained from the World Development Indicators. Column 1 contains the lights measure from Table I, and corresponds to the baseline. Column 2 contains the log of life expectancy at birth. Column 3 contains the negative of the log of total fertility. Column 4 contains the negative of the log of the number of births to mothers aged 15-19, per 100,000 mothers. Column 5 contains the negative of the log of the average number of kilograms of food by which an undernourished person falls below nutritional standards. Column 6 contains the negative log of the fraction of pregnant women who are anemic. Column 7 contains the log of the fraction of the population who declare in a household survey or census to have access to effective sanitation facilities. Column 8 contains the log of the fraction of the population who declare in a household survey or census to have access to a water source that is protected from contamination. Column 9 contains the log of the fraction of primary school-age children enrolled in primary school. Column 10 contains the log of the fraction of women who are literate. All other data definitions, inference procedures and sample selection are as in Table I.

11 Figures

Figure I (I)



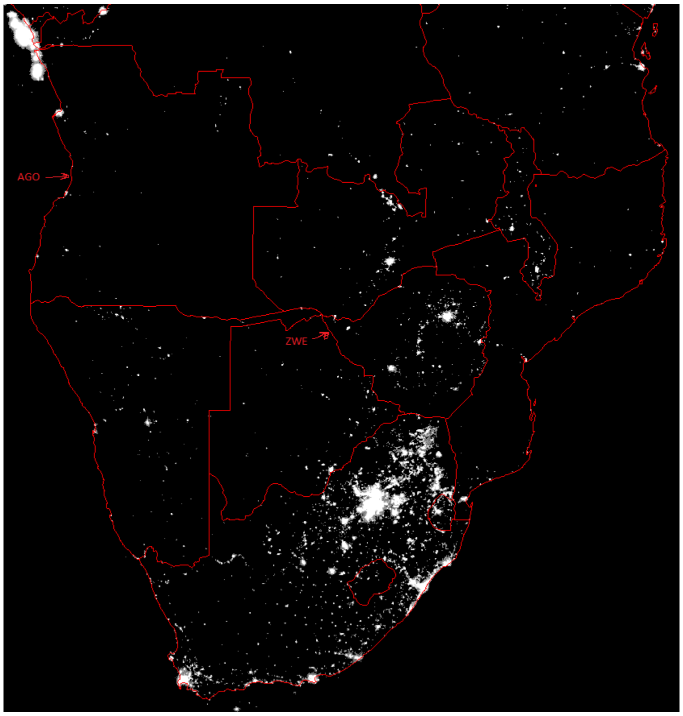
India, 1994



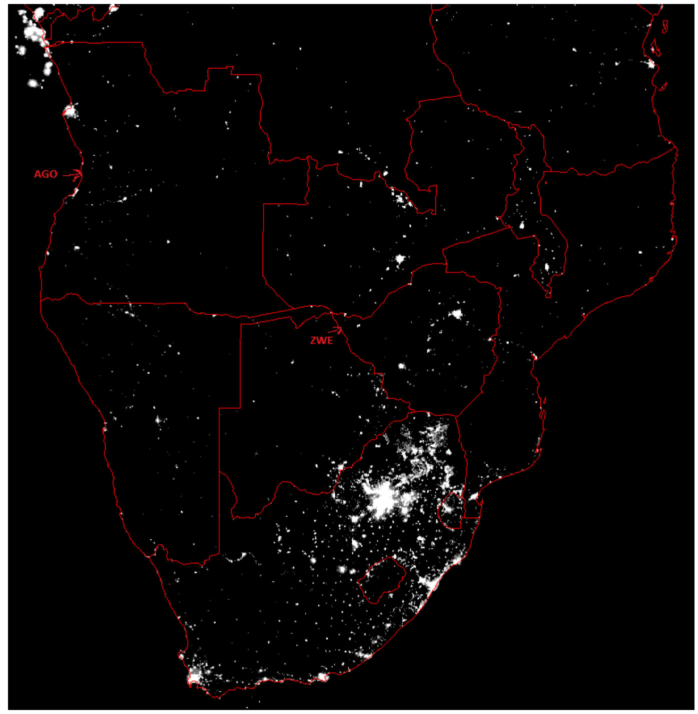
India, 2010

Data Source: NOAA.

Figure II (II)



Southern Africa, 2000



Southern Africa, 2009

Data Source: NOAA. The symbols "AGO", "ZWE" and "BWA" show Angola, Zimbabwe and Botswana respectively (the Zimbabwe symbol placed in Botswana near its Zimbabwean border to avoid masking Zimbabwean lights).

Figure III (III)

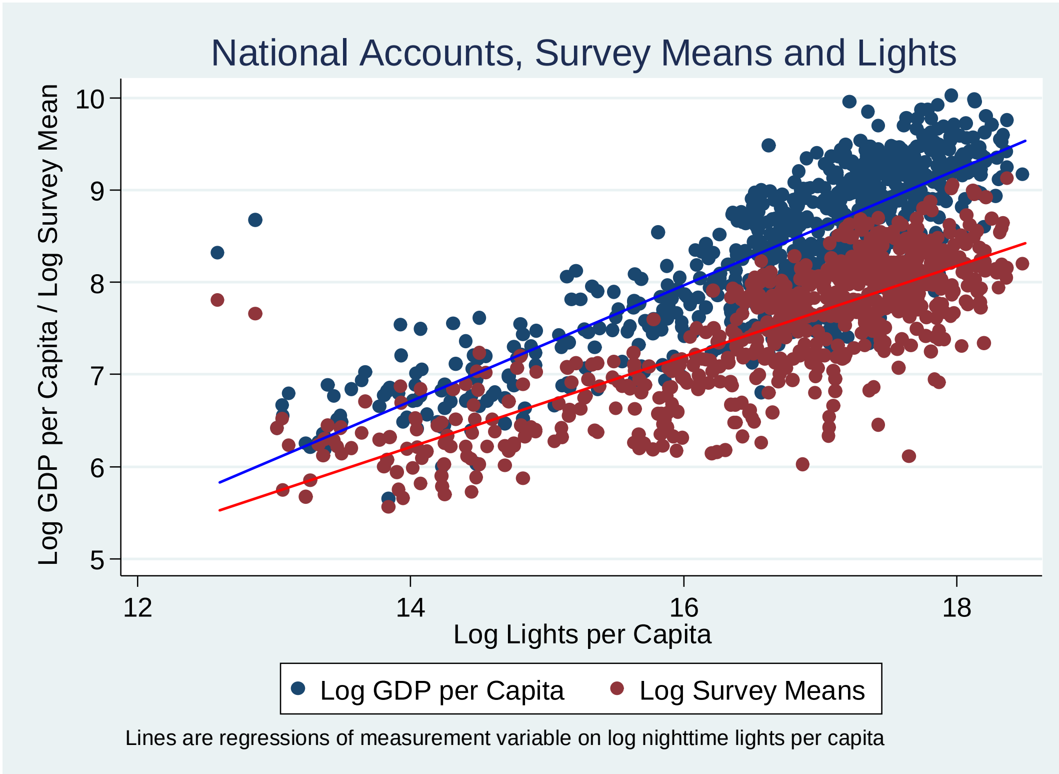
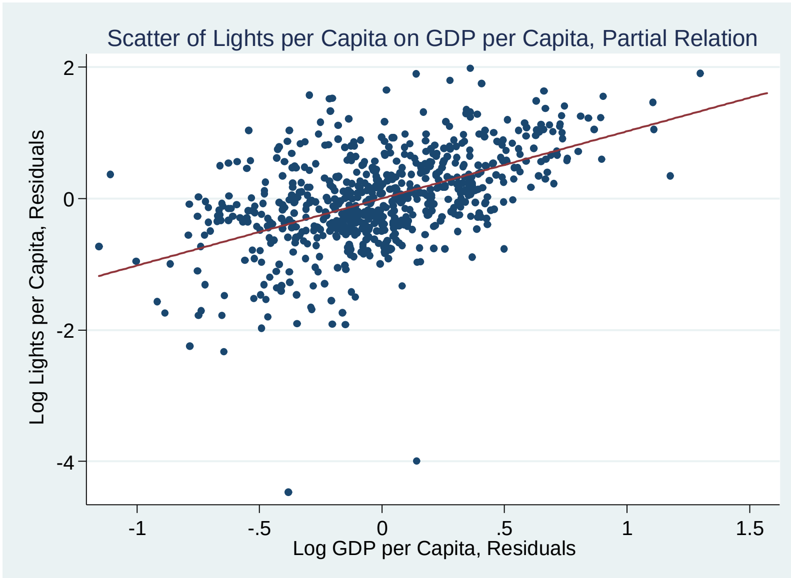
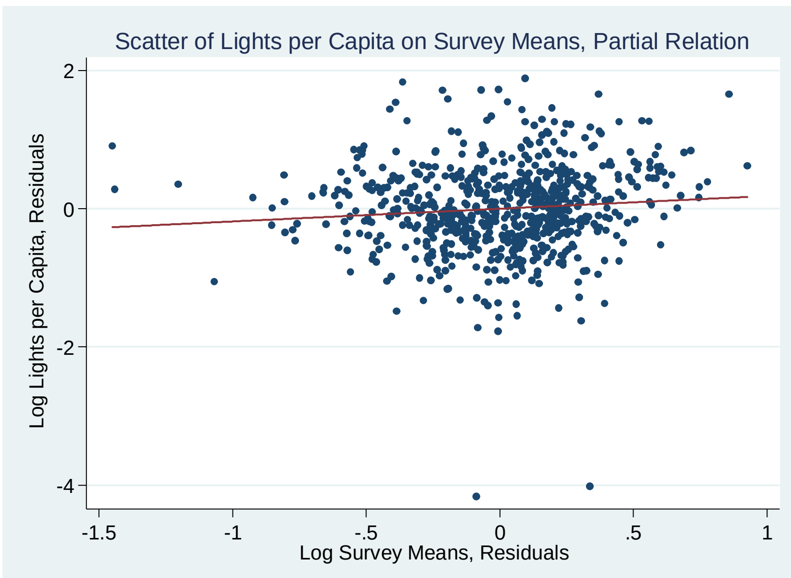


Figure IV (IV)


Scatter of Lights per Capita on GDP per Capita, Partial Relation
Log Lights per Capita, Residuals
2
0
-2
-4
-1
-.5
0
.5
1
1.5
Log GDP per Capita, Residuals


Scatter of Lights per Capita on Survey Means, Partial Relation
Log Lights per Capita, Residuals
2
0
-2
-4
-1.5
-1
-.5
0
.5
1
Log Survey Means, Residuals